```
                    GenCore version 5.1.6

          Copyright (c) 1993 - 2004  Compugen Ltd.


OM protein - protein search, using sw model

Run on:      May 7, 2004, 16:25:00 ; Search time 45 Seconds
                                      (without alignments)
                                      3814.264 Million cell updates/sec

Title:       US-10-658-989A-4
Perfect score:  3070
Sequence:    1 GSEGPBGVRGEPGPPGPAGA.........PGPSGDAGPGPPGPAGKEG 544

Scoring table:  BLOSUM62
             Gapop 10.0 , Gapext 0.5

Searched:    1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters:      1017041

Minimum DB seq length:  0
Maximum DB seq length:  2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SPTREMBL 25:*
             1:  sp_archea:*
             2:  sp_bacteria:*
             3:  sp_fungi:*
             4:  sp_human:*
             5:  sp_invertebrate:*
             6:  sp_mammal:*
             7:  sp_mhc:*
             8:  sp_organelle:*
             9:  sp_phage:*
            10:  sp_plant:*
            11:  sp_rodent:*
            12:  sp_virus:*
            13:  sp_vertebrate:*
            14:  sp_unclassified:*
            15:  sp_rvirus:*
            16:  sp_bacteriap:*
            17:  sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

```
Result         %     Query
No.   Score   Match  Length DB Length ID        Description
----  -----   -----  ------ -- ------ --------  -----------
  1   2966    96.6   1461    4  1461  O76045    O76045 homo sapien
  2   2966    96.6   1464    4  1464  Q8N473    Q8N473 homo sapien
  3   2872    93.6   1453   11  1453  Q63079    Q63079 rattus norv
  4   2857    93.1   1453   11  1450  Q8IOJ9    Q8IOJ9 mus musculu
  5   2528    82.3   1450   13  1445  Q9YIB4    Q9YIB4 cynops pyrr
  6   2525    82.2   1445   13  1449  Q93251    Q93251 rana catesb
  7   2525    82.2   1449   13  1447  Q802B5    Q802B5 xenopus lae
  8   2484.5  80.9   1447   13  1449  Q91B91    Q91B91 xenopus lae
  9   2231    72.7   1449   13  1487  Q910C0    Q910C0 oncorhynchu
 10   2225    72.5   1487    6  1160  O77753    O77753 canis famil
 11   2224    72.4   1160    4  1487  Q14046    Q14046 homo sapien
 12   2224    72.4   1487    4  1418  Q14047    Q14047 homo sapien
 13   2220    72.3   1418    6  1269  Q28396    Q28396 equus cabal
 14   2195    71.5   1269   13  1419  Q7T227    Q7T227 gallus gall
 15   2192    71.4   1419   11  1419  Q63123    Q63123 rattus norv
 16   2192    71.4   1419   11        Q80X38    Q80X38 mus musculu
 17   2192    71.4   1419   11  2192  Q80VY3    Q80VY3 mus musculu
 18   2192    71.4   1442   11  2192  Q62031    Q62031 mus musculu
 19   2192    71.4   1442   11  2192  Q62033    Q62033 mus musculu
 20   2185    71.2   1459   11  2192  Q62032    Q62032 mus musculu
 21   2147    69.9   1420   13  2185  Q90W37    Q90W37 gallus gall
 22   2145    69.9   1486   13  2147  Q91717    Q91717 xenopus lae
 23   2140    69.7   1486   13  2145  Q7ZT16    Q7ZT16 xenopus lae
 24   2128    69.3   1418   13  2140  Q9W7R9    Q9W7R9 cynops pyrr
 25   2126    69.3   1491   13  2128  Q91718    Q91718 xenopus lae
 26   2105    68.6   1491   13  2126  Q7ZTM4    Q7ZTM4 xenopus lae
 27   1953    63.6   1458   13  2105  Q910B9    Q910B9 oncorhynchu
 28   1923    62.6   1347    4  1953  Q96QB3    Q96QB3 homo sapien
 29   1915    62.4   1497   11  1923  Q7TMS0    Q7TMS0 mus musculu
 30   1910    62.2   1314   11  1915  Q8CGA5    Q8CGA5 mus musculu
 31   1899    61.9   1497   11  1910  Q61431    Q61431 mus musculu
 32   1899    61.9    998   11  1899  Q8CFM4    Q8CFM4 mus musculu
 33   1899    61.9   1222   11  1899  Q8K173    Q8K173 mus musculu
 34   1899    61.9   1464   11  1899  Q8BLW4    Q8BLW4 mus musculu
 35   1892.5  61.6   1464   11  1899  Q7TT32    Q7TT32 mus musculu
 36   1886    61.4   1366    4  1892.5 Q8BKY2   Q8BKY2 mus musculu
 37   1885    61.4   1366    4  1886  Q7Z5S6    Q7Z5S6 homo sapien
 38   1875    61.1   1352   13  1885  Q15177    Q15177 homo sapien
 39   1823    59.4   1163    4  1875  Q90YJ0    Q90YJ0 brachydanio
 40   1798    58.6   1346    4  1823  Q8N6U4    Q8N6U4 homo sapien
 41   1798    58.6   1346   13  1798  Q8UUJ3    Q8UUJ3 oncorhynchu
 42   1792.5  58.4   1346   13  1798  Q8UUJ4    Q8UUJ4 xenopus lae
 43   1635    53.3   1258   13  1792.5 Q801M5   Q801M5 xenopus lae
 44   1550.5  50.5   1414    5  1635  Q8AW11    Q8AW11 brachydanio
 45                               1550.5 Q26634 Q26634 strongyloce
```

ALIGNMENTS

RESULT 1
O76045
ID   O76045   PRELIMINARY;   PRT; 1461 AA.
AC   O76045;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Pro alpha 1(I) collagen.
GN   COL1A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85130970; PubMed=2857713;
RA   D'Alessio M., Bernard M., Pretorius P.J., de Wet W., Ramirez F.;
RT   "Complete nucleotide sequence of the region encompassing the first
RT   twenty-five exons of the human pro alpha 1 (I) collagen gene.";
RL   Gene 67:105-115(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88329734; PubMed=2843432;
RA   Chu M.L., de Wet W., Bernard M., Ramirez F.;
RT   "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT   Promoter structure, AluI repeats, and polymorphic transcripts.";
RL   J. Biol. Chem. 260:2315-2320(1985).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89025644; PubMed=3178743;
RA   Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA   Jaenisch R., Prockop D.J.;
RT   "Structure of a full-length cDNA clone for the prepro alpha 1(II) chain
RT   of human type I procollagen.";
RL   Biochem. J. 253:919-922(1988).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91138770; PubMed=1995349;
RA   Maatta A., Bornstein P., Penttinen R.P.;

17 2192   71.4   2192   11 1419   Q80VY3 mus musculu
18 2192   71.4   2192   11 1442   Q62031 mus musculu
19 2192   71.4   2192   11 1442   Q62033 mus musculu
20 2185   71.2   2192   11 1459   Q62032 mus musculu
21 2147   69.9   2185   13 1420   Q90W37 gallus gall
22 2145   69.9   2147   13 1486   Q91717 xenopus lae
23 2140   69.7   2145   13 1486   Q7ZT16 xenopus lae
24 2128   69.3   2140   13 1418   Q9W7R9 cynops pyrr
25 2126   69.3   2128   13 1491   Q91718 xenopus lae
26 2105   68.6   2126   13 1491   Q7ZTM4 xenopus lae
27 1953   63.6   2105   13 1458   Q910B9 oncorhynchu
28 1923   62.6   1953   4 1347   Q96QB3 homo sapien
29 1915   62.4   1923   11 1497   Q7TMS0 mus musculu
30 1910   62.2   1915   11 1314   Q8CGA5 mus musculu
31 1899   61.9   1910   11 1497   Q61431 mus musculu
32 1899   61.9   1899   11   998   Q8CFM4 mus musculu
33 1899   61.9   1899   11 1222   Q8K173 mus musculu
34 1899   61.9   1899   11 1464   Q8BLW4 mus musculu
35 1892.5 61.6   1899   11 1464   Q7TT32 mus musculu
36 1886   61.4   1892.5   4 1366   Q8BKY2 mus musculu
37 1885   61.4   1886   4 1366   Q7Z5S6 homo sapien
38 1875   61.1   1885   13 1352   Q15177 homo sapien
39 1823   59.4   1875   4 1163   Q90YJ0 brachydanio
40 1798   58.6   1823   4 1346   Q8N6U4 homo sapien
41 1798   58.6   1798   13 1346   Q8UUJ3 oncorhynchu
42 1792.5 58.4   1798   13 1346   Q8UUJ4 xenopus lae
43 1635   53.3   1792.5 13 1258   Q801M5 xenopus lae
44 1550.5 50.5   1635   13 1414   Q8AW11 brachydanio
45   50.5   1550.5   5   Q26634 strongyloce

RT "Highly conserved sequences in the 3'-untranslated region of the
RT COL1A1 gene bind cell-specific nuclear proteins.";
RL FEBS Lett. 279:9-13(1991).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=92157916; PubMed=1787829;
RA Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
RA Olsen A.S., Prockop D.J.;
RT "Completion of the last half of the structure of the human gene for
RT the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
RL Matrix 11:375-379(1991).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=98107942; PubMed=9443882;
RA Korkko J., Ala-Kokko L., De Paepe A., Nuytinck L., Earley J.,
RA Prockop D.J.;
RT "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and
RT scanning by conformation-sensitive gel electrophoresis identifies only
RT COL1A1 mutations in 15 patients with osteogenesis imperfecta type I:
RT identification of common sequences of null-allele mutations.";
RL Am. J. Hum. Genet. 62:98-110(1998).
RN [7]
RP SEQUENCE FROM N.A.
RA Korkko J.M., Earley J.J., Nuytinck L., DePaepe A., Prockop D.J.,
RA Ala-Kokko L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF017178; AAB94054.2; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS50184; VWFC_2; 1.
KW Collagen.
SQ SEQUENCE 1461 AA; 138630 MW; 9ACF6DE30EA78E21 CRC64;

Query Match 96.6%; Score 2966; DB 4; Length 1461;
Best Local Similarity 95.6%; Pred. No. 1.3e-169;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GSEGPEGVRGEPGPPGPAGAAGPAGDPAGDGEPGAKGADGAPGIAGAPGFPGARGPSGPE 60
GSEGPEGVRGEPGPPGPAGAAGPAGDPAGDGEPGAKGADGAPGIAGAPGFPGARGPSGPE
Db 359 GSEGQGGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ 418

Qy 61 GPGGPPGPGPKGDSGPEGPAPGSKGDTGAKGEPGPVGEGKPGARGEPGPTGLP 120
GPGGPPGPGPKGDSGPEGPAPGSKGDTGAKGEPGPVGEGKPGARGEPGPTGLP
Db 419 GPGGPPGPGPKGNSGPEGPAPGSKGDTGAKGEPGPVGQGPPGPGARGEGRPGPTGLP 478

Qy 121 GPPGERGGPGSRGFPGPADGVAGPKGPAGERGSPGEAGRPGEAGLPGAKGLT 180
GPPGERGGPGSRGFPGPADGVAGPKGPAGERGSPGEAGRPGEAGLPGAKGLT
Db 479 GPPGERGGPGSRGFPGPADGVAGPKGSPGPAGERGSPGEAGRPGEAGLPGAKGLT 538

Qy 181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGARGEAGVMGFPGPKGAAGEPGKAGERGVP 240
GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGARGEAGVMGFPGPKGAAGEPGKAGERGVP
Db 539 GSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGVP 598

Qy 241 GPPGAVGPAGKDGEAGAGPPGPAGERGEEGPAGSPCFEGLPGPAGPPGEAGKPGE 300
GPPGAVGPAGKDGEAGAGPPGPAGERGEEGPAGSPCFEGLPGPAGPPGEAGKPGE
Db 599 GPPGAVGPAGKDGEAGAQGPPGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQ 658

Qy 301 GVPGDLGAPGPSGARGEPGPPGERGVQGPPGPAGPRGPANGAPGNDAGKGDGAPAGPGSQ 360
GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGPANGAPGNDAGKGDGAPAGPGSQ
Db 659 GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGPANGAPGNDAGKGDGAPAGPGSQ 718

Qy 361 GAPGLEGMPGERGAAGLPGPKGDRDGAPGPKGDVRGLTGPIGPPGPAGAPGDK 420
GAPGL:GMPGERGAAGLPGPKGDR:GAPGPKGDVRGLTGPIGPPGPAGAPGDK
Db 719 GAPGLQMPGERGAAGLPGPKGDRGDAGPKGDVRGLTGPIGPPGPAGAPGDK 778

Qy 421 GESGPSGPAGPTGARGAPGDRGEPGPGPGPAGFAGPPGPADGEPGAKGEPGDAGAKGDAGAPGP 480
GESGPSGPAGPTGARGAPGDRGEPGPGPGPAGFAGPPG:::::::::
Db 779 GESGPSGPAGPTGARGAPGDRGEPGPPGPGPAGFAGPPGPADGQPGAKGEPGDAGAKGDAGPP 838

Qy 481 GPAGPAGPPGIDVGAPGAKGARGSAGPPGATGPGAAGRVGPGPSGDAGPPGPPGPA 540
GPAGPAGPPGI:::::::::::::::::::::::::::::
Db 839 GPAGPAGPPGIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPGPGPSGNAGPPGPPGPA 898

Qy 541 GKEG 544
GKEG
Db 899 GKEG 902

RESULT 2
Q8N473
ID Q8N473 PRELIMINARY; PRT; 1464 AA.
AC Q8N473;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC036531; AAH36531.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS50184; VWFC_2; 1.
KW Hypothetical protein; Collagen.
SQ SEQUENCE 1464 AA; 139011 MW; B0581F8D1C89DDE8 CRC64;

Query Match 96.6%; Score 2966; DB 4; Length 1464;
Best Local Similarity 95.6%; Pred. No. 1.3e-169;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GSEGPEGVRGEPGPPGPAGAAGPAGDGEPGAKGADGAPGFPGARGPSGPE 60
GSEGP:GVRGEPGPPGPAGAAGPAG:PAGD:GEPGAKGA:GAPGIAGAPGFPGARGPSGPQ
Db 362 GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ 421

Qy 61 GPGGPPGPGPKGDSGPEGPAPGSKGDTGAKGEPGPVGEGKPGARGEPGPTGLP 120
GPGGPPGPGPKG:SGPEGPAPGSKGDTGAKGEPGPVG:G:PG:RGE:GPTGLP
Db 422 GPGGPPGPGPKGNSGPEGPAPGSKGDTGAKGEPGPVGQGPPGPGARGEGRPGPTGLP 481

Qy 121 GPPGERGGPGSRGFPGPADGVAGPKGPAGERGSPGEAGRPGEAGLPGAKGLT 180
GPPGERGGPGSRGFPGPADGVAGPKG:PG:PAGERGSPGEAGRPGEAGLPGAKGLT
Db 482 GPPGERGGPGSRGFPGADGVAGPKGSPGPAGERGSPGEAGRPGEAGLPGAKGLT 541

Qy 181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGARGEAGVMGFPGPKGAAGEPGSE 240
GSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGVP
Db 542 GSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGVP 601

```
Qy     1  GSEGPEGVRGEPGPPGPPGPAGAPGPADGEPGPAGAAGPAGDGAPGIAGAPGFPGARGPSGPE   60
          |||||::||||||||||||||||| |||||||| ::|||||||| |:|||||||||||||||
Db   351  GSEGPQGVRGEPGPPGPPGPAGAAGPAGNPGADGQPGAGENGAPGIAGAPGFPGARGPSGPQ  410

Qy    61  GPGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEEGKPGARGEPGPGPTGLP  120
          || ||||| ||||||| ||:|||||||||| :||||||||||| |:|||||||||| |||
Db   411  GPSAGAPGKTSGEPGAPGDPNKGDTGAKGEPGPAGVQGPPGPAGEGKRGARGEPGPSGLP  470

Qy   121  GPPGERGGPGSRGPGADGVAGPKGSPGAGERGSPGPAGPKGSPGAGRPGEAGLPGAKGLT  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   471  GPPGERGGPGSRGPGADGVAGPKGSPGAGERGSPGPAGPKGSPGAGRPGEAGLPGAKGLT  530

Qy   181  GSPGSPGPDGKTGPPGPPGPAGEDRPGPPGPPGPAGRGEAGVMGPPGPKGAAGEPGKAGERGVP  240
          |||||||||||||||||||||| |:|||  |||||||| |||:||:||||| ||||||||||||
Db   531  GSPGSPGPDGKTGPPGPPGPAGQDRGPGPGPAGRGEAGQAGVMGFPGPKGTAGEPGKAGERGVP  590

Qy   241  GPPGAVGPVAGKDGEAGAEGPGPPGPAGERGEEGPAGSPGFEGLPGPAGPPGEAGKPGEE  300
          |||||||| |||||||||:|| |||||||||| |||||||||:|||||||||||||||||
Db   591  GPPGAVGPVAGKDGEAGAQGPGPPGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQ  650

Qy   301  GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPGDAGPGDDGAKGDAGAPGAPGSE  360
          |||||||||||||||||:||||||||:|||||||||||||||||||||||||||||:
Db   651  GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPGDAGPGDDGAKGDAGAPGAPGSQ  710

Qy   361  GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADSPGKDGVRGLTGPIGPPGPAGPGDK  420
          ||||| ||:|||||||||||||:|||||||||||:|||||||||||||||||||||||
Db   711  GAPGLQGMPGERGAAGLPGPKGDKGDRGDAGPKGADGPGKDGVRGLTGPIGPPGPAGPGDK  770

Qy   421  GESGPSGPAGTGARGAGPPGATGFPGAAGKGARGSAGPPGATGFPGAAGRVGPPGSGDAGPPGPGPA  480
          ||:||:|||||||||  |||||||||||| ||||||||||||||||||||||||||| |||||||||
Db   771  GEAGPSGPAGTGARGAPGDRGEAGPPGPGADGPPGVKGDAGPP  830

Qy   481  GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGSGDAGPGPPGPA  540
          ||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||
Db   831  GPAGPAGPPGPIGNVGAPGPKGSRGAAGPPGATGFPGAAGRVGPPGSGNAGPPGPGPV  890

Qy   541  GKEG  544
          ||||
Db   891  GKEG  894
```

Query Match   93.6%; Score 2872; DB 11; Length 1453;
Best Local Similarity   92.1%; Pred. No. 5.2e-164;
Matches 501; Conservative 26; Mismatches 17; Indels 0; Gaps 0;

Left column (RESULT 5)

```
SQ   SEQUENCE  1453 AA;  138032 MW;  0B7F06BBB9A1D5EA CRC64;

Query Match             93.1%;  Score 2857;  DB 11;  Length 1453;
Best Local Similarity   91.4%;  Pred. No. 4.1e-163;
Matches  497;  Conservative  28;  Mismatches  19;  Indels  0;  Gaps  0;

Qy      1 GSEGPEGVRGEPGPPGPAGAAGPAGDDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE  60
          ||||||||||||| |:||::|| ||||||||:||::||||||||||||||||||||||||
Db    351 GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ  410

Qy     61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEGKPGARGEPGPTGLP  120
          ||||||||||| ||||:|||||| ||||||||||||||||||||||::||||||||||
Db    411 GPSGPPGPKGNSGEPGAPGNKGDTGAKGEPGATGVQGPPGPAGEGKGRARGEPGPSGLP  470

Qy    121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEARGRPGEAGLPGAKGLT  180
          ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db    471 GPPGERGGPGSRGFPGADGVAGPKGSGERGAPGPAGPKGSPGEARGRPGEAGLPGAKGLT  530

Qy    181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGARGEAGVMGFPGPKGAAGEPGKAGERGVP  240
          ||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db    531 GSPGSPDGKTGPPGPAGQDGRPGPAGPPGARGEAGVMGFPGPKGTAGEPGKAGERGLP  590

Qy    241 GPPGAVGPAGKDGEAGAEGPPGPAGERGEEGPAGSPGFEGLPGAPGPGEAGKPGEE  300
          ||||||||||||||||||:|||||||||||||||:||||:||||||||||||||||
Db    591 GPPGAVGPAGKDGEAGAQGAPGPAGERGEQGPAGSPGFQGLPGAPGPGEAGKPGEQ  650

Qy    301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGADGAPGDDGAKGDAGAPGAPGSE  360
          ||||||||||||||||||||||||:|||||||||||| :|||||||||||||||||||
Db    651 GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGNNGAPGNDGAKGDTGAPGAPGSQ  710

Qy    361 GAPGLEGMPGERGAAGLPGPKGRDGAPGKDGVRGLTGPIGPPGPAGAPGDK  420
          |||||||||||||||||||||||||| :||||||||||||||||||||||||
Db    711 GAPGLQGMPGERGAAGLPGPKGDGDAPGKGADGSPGKDGARGLTGPIGPPGPAGAPGDK  770

Qy    421 GESGPSGPAGPTGARGAPGPTGARGSPGEARGRPGEAGPPGATGFPGAAGKDGAGPP  480
          |||||||||||||||||||:|||||||||||||||||||||||||:||||||||||||
Db    771 GEAGPSGPPGPTGARGEAGPPGPAGPPGATGFPGAAGKEPGDTGVKGDAGPP  830

Qy    481 GPPGPAGPPGPIGDVGAPGAKGARGSAGPPGDATGFPGAAGRVGPPGPSGDAGPPGPPGPA  540
          ||||||||||||:||||||||||||:||||||||||||||||||||||||||||||:
Db    831 GPPGPAGPPGPIGNVGAPGPKGPRGAAGPPGAAGRVGPPGPSGNAGPPGPPGPGPV  890

Qy    541 GKEG  544
          ||||
Db    891 GKEG  894

RESULT 5
Q9YIB4
ID   Q9YIB4                   PRELIMINARY;       PRT;  1450 AA.
AC   Q9YIB4;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Alpha 1 type I collagen.
OS   Cynops pyrrhogaster (Japanese common newt).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.
OX   NCBI_TaxID=8330;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Regenerate forelimbs;
RX   MEDLINE=99407244; PubMed=10474166;
RA   Asahina K., Obara M., Yoshizato K.;
RT   "Expression of genes of type I and type II collagen in the formation
RT   and development of the blastema of regenerating newt limb.";
RL   Dev. Dyn. 216:59-71(1999).
EMBL; AB015438; BAA36973.1; -.
DR   GO; GO:0005581; C:collagen; IEA.
DR   GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
```

Right column (RESULT 6)

```
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD000007; Clg_helix; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
KW   Collagen.
SQ   SEQUENCE  1450 AA;  137564 MW;  ABF8A74841B87B7C CRC64;

Query Match             82.3%;  Score 2528;  DB 13;  Length 1450;
Best Local Similarity   80.3%;  Pred. No. 1.8e-143;
Matches  437;  Conservative  43;  Mismatches  64;  Indels  0;  Gaps  0;

Qy      1 GSEGPEGVRGEPGPPGPAGAAGPAGDGPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE  60
          |||||||||||||:||::|| |||| |||||||:|||||||||||||||||||||||||
Db    348 GSEGPQGVRGEPGPPGPAGAAGPSGNPGTDGQPGAKGATGSPGIAGAPGFPGARGPSGPQ  407

Qy     61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEGKPGARGEPGPTGLP  120
          ||||||||||::|||||||:||:|||||||||||||||||||||||:||||||||||
Db    408 GPAGAPGPKGNNGRPGAQGNKGEPGAKGEPGPGPVGQGPPGPSGEGKRGSRGEPGPAGPP  467

Qy    121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEARGRPGEAGLPGAKGLT  180
          ||||| | |||||||||:||:||||||||::||||||||| ||||||||||||||||||
Db    468 GPAGERGGPGSRGFPGSDGASGPKGAPGERGSVGPAGPKGSTGESGRFGEPGLPGAKGLT  527

Qy    181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGARGEAGVMGFPGPKGAAGEPGKAGERGVA  240
          ||||||||||||||||||||| :||||||||||:||||||||||||||||||||||||
Db    528 GSPGSPGPDGKTGPAGAAGQDGHPGPPGPPGSGARQSGVNGFPGPKGAAGEPGKSGERGVA  587

Qy    241 GPPGAVGPAGKDGEAGAEGPPGPAGPPGPGEGLPGPAGSPGFQGLPGSPGPGEAGKPGEE  300
          ||||||||||||||||||:|||||||||||||:|||||| ::|| :||:||||||||||
Db    588 GPPGATGAPGKDGEAGAQGPPGSGSGEQGPAGSPGFQGLPGPGEAGAGSPGPGEAGKPGEQ  647

Qy    301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGADGAPGDDGAKGDAGAPGAPGSE  360
          |||||:|||||||||||||:||||:|:|:|||||||||||||||||||||||||||||
Db    648 GAPGDAGGPGPSGPRGERGFPGERGGQGPGAGPAGAQGPRGSPSGPGNDGAKGEAGAAGPGGR  707

Qy    361 GAPGLEGMPGERGAAGLPGPKGRDGAPGKDGVRGLTGPIGPPGPAGAPGDK  420
          |||||||||||||:|||||||| ||||||||:||:|||||||||||:||||||
Db    708 GPPGLQGMPGERGSAGMPGAKGDRDGAGTKGADGAPGKDGARGLTGPIGPPGSGAPGDK  767

Qy    421 GESGPSGPAGPTGARGAPGPTGARGSPGEARGRPGEAGPPGATGFPGAAGKDGAGPP  480
          |||||||||||||||||||||||||||||||||||||||||||||||||::|||||
Db    768 GEGGPSGPSGPTGARGSPGEARGRPGEAGPPGAGPGDRGEPGPPGPAGFAGPGAGPPGPAGP  827

Qy    481 GPPGPAGPPGPIGDVGAPGAKGARGSAGPPGDATGFPGATGFPGPGPSGDAGPPGPPGPGPA  540
          ||:||||||||||:||||||||| |||||||||||:|||||||:
Db    828 GPAGPTGAPGPAGNVGAPGPKGTRGAAGPPGATGFPGAAGRLGPPGAAGPPGPPGPGPG  887

Qy    541 GKEG  544
          ||||
Db    888 GKEG  891

RESULT 6
093251
ID   093251                   PRELIMINARY;       PRT;  1445 AA.
AC   093251;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Alpha 1 type I collagen.
OS   Rana catesbeiana (Bull frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Ranidae; Rana.
OX   NCBI_TaxID=8400;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99294154; PubMed=10367734;
RA   Asahina K., Utoh R., Obara M., Yoshizato K.;
```

```
RT   "Cell-type specific and thyroid hormone-dependent expression of genes
RT   of a1(I) and a2(I) collagen in intestine during
RT   amphibianmetamorphosis.";
RL   Matrix Biol. 18:89-103(1999).
DR   EMBL; AB015440; BAA29028.1; -.
DR   GO; GO:0005581; C:collagen; IEA.
DR   GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR   InterPro; IPR008161; C1q_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR002181; Fibrinogen_C.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD000007; C1q_helix; 3.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
KW   Collagen.
SQ   SEQUENCE  1445 AA;  137252 MW;  F59BB550C9873F04 CRC64;

Query Match            82.2%;  Score 2525;  DB 13;  Length 1445;
Best Local Similarity  81.1%;  Pred. No. 2.8e-143;
Matches  441;  Conservative  37;  Mismatches  66;  Indels  0;  Gaps  0;

Qy    1  GSEGPEGVRGEPGPPGPAGAAGPAGDGPADGEPGAKGADGAPGIAGAGFPGARGPSGPE   60
            : :  ||| |   :||  :|| ||| : |  |||||| :||||||||||||||  ::
Db  347  GPDGPQGGARGEPGAPGQAGPAGSAGNPGTDGQPGAKGATGAPGIAGAGFPGARGAPGPQ  406

Qy   61  GPGGPPGPKGDSGEPGAPGSPGPKGNNGEPGAQGNKGEPGAGPGPG...EGKRGSRGEPGPSGPP  120
           ||| ||||| :||  :|
Db  407  GPGGSPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEGKRGSRGEPGPSGPP  466

Qy  121  GPGERGGPGSRGFPGADGAGGRPGPAGEDGRPGPPGPGAARGEAGVMGFPGPKGAAGEPGKAGERGVP  180
Db  467  GPAGERGAPGSRGFPGADGAGGRPGPAGEDGRPGPPGPGARGEAGVMGFPGPKGAAGEPGKAGERGVA  526

Qy  181  GSPGSPGPDGKTGPPGPAGEDGRPGPPGPGARGEAGREEGPAGSPGFEGLPGPAGPPGEAGKPGEE  240
Db  527  GSPGSPGPDGKTGPAGAPGDGRPGPPGPGARGQSGVMGFPGPKGAAGEPGKAGERGERGVA  586

Qy  241  GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEEGPAGSPGFEGLPGPAGPPGEAGKPGEE  300
Db  587  GPPGAVGPAGKDGEAGAQGGPPGPAGPGGQDGRAGPPGPPGPAGEPGKNGEKGVA  646

Qy  301  GVPGDLGAPGPSGARGEPGPPGERGVEGPPGPAGPPGEDGAKGDAGAPGAPGSE  360
Db  647  GAPGDVGPSGPAGSRGERGFPGERGAIGPPGPQGPPGQGPPGARGQSGVMGFPGPKGAAGEPGKNGEKGVA  706

Qy  361  GAPGLBGMPGBRGAAGLPGPKGDRDGAGLPGPKGDRDQGGPKGADGPGDVRGLTGPIGPPGPAGPGDK  420
Db  707  GVPDGVGPSGPAGSRGERGFPGERGAQGPPGQGPAGRGNGAPGNDGAKGEAGAPGAPGGQ  766

Qy  421  GEAGPAGPAGPTGSRGAPGERGEPGSGPAGEGPGPGDAGPKGEPGDAKGENGGPKGADGPAGPP  480
Db  767  GEAGPAGPAGPTGSRGAPGERGEPGSGPAGEGPGPGDAGPKGEPGDAKGEQGDAGPPGPGPPGPA  826

Qy  481  GEAGPAGPGPIGDVGAPGAKGARGSAGPGPCATGPGEAAGRVGPPGPSGDAGPPGPPGPPGPA  540
Db  827  GAAGPTGAPGPAGALGSPGPKGARGAPGFPGAAGRLGPPGPSGNAGPPGPSGPA  886

Qy  541  GKEG  544
         ||||
Db  887  GKEG  890

RESULT
ID   Q802B5        PRELIMINARY;      PRT;  1449 AA.
AC   Q802B5;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
```

```
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   Klein S., Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC049829; AAH49829.1; -.
DR   GO; GO:0005581; C:collagen; IEA.
DR   GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE  1449 AA;  137464 MW;  4B32E20BECCE978C CRC64;

Query Match            82.2%;  Score 2525;  DB 13;  Length 1449;
Best Local Similarity  80.3%;  Pred. No. 2.8e-143;
Matches  437;  Conservative  50;  Mismatches  57;  Indels  0;  Gaps  0;

Qy    1  GSEGPEGVRGEPGPPGPAGAAGPAGDGPADGEPGAKGADGAPGIAGAGFPGARGPSGPE   60
Db  348  GSDGPQGGRGGEPGAPGQAGAAGPSGNPGSDGQPGAKGATGAPGIAGAGFPGARGAPGAQ  407

Qy   61  GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEGKPGARGEPGPTGLP  120
Db  408  GPGGSPGPKGNNGEPGAQGNKGEAGAGPGPAGPGPAGVQGPPGPSGEEGKRGSRGEPGPAGPP  467

Qy  121  GPGERGGPGSRGFPGADGAGGRPGPPGPGADGVAGPKGPAGERGSPGAPGKGSPGEAGRPGEAGL PGAKGLT  180
Db  468  GPAGERGGPGSRGFPGSDGASGPKAGAGEPGPVGPAGRGRGPVGPAGGRGPVGPKGSGSGRPGBPGLPGAKGLT  527

Qy  181  GSPGSPGPDGKTGPPGPAGEDGRPGPPGPGARGQSGVMGFPGPKGAAGEPGKAGERGVA  240
Db  528  GSPGSPGDGKTGPPGPAGEDGRGRPGPPGPAGRGRGEAGVMGFPGPKGAAGEPGKAGERGVP  587

Qy  241  GPPGAVGPAGKDGEAGAQGGPPGPAGPAGRGREGEQGPAGPGFQGLPGSPGAPGESGKPGEQ  300
Db  588  GPPGAVGLPGKDGDAGAQGPPGPAGPAGERGEQGPAGPGFQGLPGPPGPAGESGKPGEQ  647

Qy  301  GVPDGLGAPGPSGARGEPGFPGERGVEGPPGPGERGVEGPPGPGADGAPGASGE  360
Db  648  GVPGDVGPSGPAGARGERGFPGERGAQGGPPGPGARGSNGAPGNDGAKGEAGAAGAPGGQ  707

Qy  361  GAPGLEGMPGBRGAAGLPGPKGDRGDRAGPPGPAGKGEPGDAKGKDAGPP  420
Db  708  GAPGLQGMPGERGSSGLPGAKGDRGDQGVKGSDGTPGKDGVRGLTGPIGPPGPAGPGDK  767

Qy  421  GESGPSGPAGPTGARGAPGTGARGAPGRGESGPPGPAGPAGPAGPGADGQPGADGGQPGADGGP  480
Db  768  GEAGPAGPAGPTGARGAPGERGESGPPGPAGPAGPGADGQPGADGAKGDAGPP  827

Qy  481  GPAGPAGPGPIGDVGAPGAKGARGSAGPPGPATGFPGAAGRVGPPGPSGDAGPPGPPGPA  540
Db  828  GPAGPTGAPGPAGALGSPGPKGARGAPGFPGATGFPGAAGRLGPPGPSGNAGPPGPSGPA  887

Qy  541  GKEG  544
Db  888  GKEG  891
```

RESULT 8
Q9IB91
ID   Q9IB91   PRELIMINARY;   PRT;   1447 AA.
AC   Q9IB91;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Type I collagen alpha 1.
GN   COL1A1.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Goto T., Katada T., Kinoshita T., Kubota H.Y.;
RT   "Expression and characterization of Xenopus type I collagen alpha 1
RT   (COL1A1) during embryonic development.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB034701; BAA94972.1; -.
DR   GO; GO:0005581; C:collagen; IEA.
DR   GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD000007; Clg_helix; 2.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
KW   Collagen.
SQ   SEQUENCE   1447 AA; 137446 MW; A4A6DD2B4158E38B CRC64;

Query Match   80.9%; Score 2484.5; DB 13; Length 1447;
Best Local Similarity   79.6%; Pred. No. 7.3e-141;
Matches 433; Conservative 50; Mismatches 60; Indels 1; Gaps 1;

Qy   1 GSEGPEGVRGEPGPPGPAGAAGPEPGDGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE 60
           ||||||:|::|||||||| |:|:||||||| |||||:|||||:||
Db   348 GSDGPQGSRGEPGEPGAPGQAGAAGPGNPGSDGQPGAKGATGAPGIAGAPGFPGARGAPGAQ 407

Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGAGEEGKPGARGEPGPTGLP 120
           || ||||| |||:||||||||| |:|||||||||||||:|||||:|||||:||||||||
Db   408 GPGGSPGPGKGNNGEPGAQGNKGEAGAKGEPGPAGVQGPPGPGEEGKRGSRGEP-PAGPP 466

Qy   121 GPPGERGGPGSSRGFPGDADGVAGRPKGPAGERGSPGPAGRGSPGEAGRPGEAGLPGAKGLT 180
           || |:|||||:|||||||||:||:|||||||||| |||||||||:|:|||||:||||||
Db   467 GPAGERGGPGSKRGFPGDSDGASGPKGAPGERGVPGAPGERGVPGPGARGEAGVMGFPGPKGKGAAGEPGKAGERGVP 526

Qy   181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPGARGEAGVMGFPGPKGKGAAGEPGKAGERGVP 240
           ||||:|||:|||:|||:|||:| |:|: ||||||||:||||||||:||:|:||
Db   527 GSPGSPGSDGKTGPAGPAGPQGQDGRAGPGPGPAGPGPAGERGEGPAGSPGFEGLPGPAGPGFEAGKPGEE 586

Qy   241 GPPGAVGPAGKDGEAGAEGPPPGPAGPAGERGEGPAGSPGFEGLPGPAGPGEAGKPGEE 300
           |||||:| ||:||||||||||||||||||||||||||||:|||:||||||:|||||||
Db   587 GPPGAVGLPGKDGDAGAQGPPGPGEQGPGRGEGPAGEQGPFQGLPGSPGPAGESGKPGEQ 646

Qy   301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGPGPDGDAKGDAGAPGAPGSE 360
           |:|||:|||||| | ::|||||||:|:||||||| |||||||::||:|::|||:||
Db   647 GVPGDVGPSGPAGARGERGFPGERGAQGPPGPGPGVKGDGVRGLTGPIGPGPAGAPGGQ 706

Qy   361 GAPGLEGMPGERGAAGLPGPKGRDGAGPKGADGSPGKDGVRGLTGPIGPGPAGAPGDK 420
           |||||: ||||||||||||:|||||:||||| | |:|::|||||||||||:||||||
Db   707 GSPLQGMPGERGSSGLPGAKGDRGDQGDVKGSDGTPGADGVRGLTGPIGPPGGGAPGDK 766

Qy   421 GESGPSGPAGPTGARGAPGDRGEPGEPGPPGPAGPFAGPPGADGEPGAKGEPGDAGAKGDAGAPP 480
           ||||:||||||:|||:||||||||||||||

RESULT 9
Q910C0
ID   Q910C0   PRELIMINARY;   PRT;   1449 AA.
AC   Q910C0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Collagen al(I).
GN   COL1A1.
OS   Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC   Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX   NCBI_TaxID=8022;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21257802; PubMed=11358497;
RA   Saito M., Takenouchi Y., Kunisaki N., Kimura S.;
RT   "Complete primary structure of rainbow trout type I collagen
RT   consisting of a1(I)a2(I)a3(I) heterotrimers.";
RL   Eur. J. Biochem. 268:2817-2827(2001).
DR   EMBL; AB052835; BAB55661.1; -.
DR   GO; GO:0005581; C:collagen; IEA.
DR   GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR002181; Fibrinogen_C.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD000007; Clg_helix; 2.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
KW   Collagen.
SQ   SEQUENCE   1449 AA; 137117 MW; 62EEF8A7BFD652B8 CRC64;

Query Match   72.7%; Score 2231; DB 13; Length 1449;
Best Local Similarity   73.2%; Pred. No. 1e-125;
Matches 398; Conservative 42; Mismatches 104; Indels 0; Gaps 0;

Qy   1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE 60
           ||||||||||||||||||| |:||||||||||||||||||:|:||||||||||||:
Db   347 GGEGPQGSRGEAGNFGPAGAAGPAGNNGADGNPGTKGAPGSSGIAGAPGFPGARGPPGPGPG 406

Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGAGEEGKPGARGEPGPTGLP 120
           ||||||||||||||||||||:| |:|||||:||||||:|||||:||||||:|||||||
Db   407 GAGGAPGPKGNTGEVGANGAKGEAGAKGESGPAGVQGPAGPAGEGKRGGRGEPGAGAGAR 466

Qy   121 GPPGERGGPGSSRGFPGDADGVAGRPKGPAGERGSPGPAGRGSPGEAGRPGEAGLPGAKGLT 180
           |||||||||||||:||||||||:||:||||||||||:|||||||||||||:||:||||||
Db   467 GAPGERGAPGSRGFPGDSDGASGPKGGPGERGAGVAGAKGNTGEPGRNGEPGMPGSKGMT 526

Qy   181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPGADGEPGAGAGVMGFPGPKGKGAAGERGVP 240
           ||||||||||||||||||||:||||:||||||||||||||:||||||||:||:||||||
Db   527 GSPGSPGPDGKTGPSGAGGQDGRPGPPGPVGARGQPGVMGFPGPKGAAGEGKGPGARGERGVM 586

```
Qy  241  GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEBGPAGSPGFEGLPGAGPPGEAGKPGEE  300
              ||||: ::||||| ||  | |  ||  |   || | | |||| ||  |  |||| :
Db  587  GPSGAVGAPGKDGDGDVGAPGAPGVVAGPSGERGEVGPPGERGVEGPPGPAGPDGAKGDAGARGAPGAGPSE  646

Qy  301  GVPGDLGAPGSGARGEPGPGERGVEGPPGPAGPAGDGAKGDAGARGAPGAPGSE  360
              |::||:|| ||:||||| ||   |||  ||||   ||   ||  |||  |||||
Db  647  GLPGEGAPGSAGPSGRGDRGFPGERGAPGSGPAGARGSPGSAGNDGGKGEAGAAGAPGAGPDGQ  706

Qy  361  GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  420
              || |  |  |||:|| | | | |  | | | | ||| |  |  |  || || | |||
Db  707  GPPGLQGMPGERGAGGLPGLKGDRGDQGVKGADGAGGGKDGVRGMTGPIGPNGPAGSPGDK  766

Qy  421  GESGPSGPAGPTGARGAPGDRGERGPPGPAGEDGRPGPAGEDGPGAKGEPGDAGAKGDAGPP  480
              ||:||||| | ||| | ||  || ||| |||  |||  | | | ||   ||  ||:: ||
Db  767  GETGAPGAVGPSGARGAPGERGESGAPGPAGPAGPAGPPGGDGQPGAKGEAGDNGAKGDGGAQ  826

Qy  481  GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGDATGFPGAAGRVGPPGSGDAGPPGPPGPA  540
              |  |  |  | |  |  |||||:  |||||:||| | |  || |||| |  ||||||| ||
Db  827  GPAGPGTGAPGPQGPAGNTGAKGARGAAGPPGDATGFPGAAGRFGPPGPGPSGNNGPPGTPGPG  886

Qy  541  GKEG  544
              ||||
Db  887  GKEG  890
```

```
RESULT 10
O77753
ID   O77753        PRELIMINARY;       PRT;   1487 AA.
AC   O77753;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Type IIA procollagen.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98340920; PubMed=9676231;
RA   Du F., Acland G.M., Ray J.;
RT   "Differential splicing of type II procollagen mRNA in canine retina.";
RL   Anim. Biotechnol. 9:15-20(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20480698; PubMed=11024291;
RA   Du F., Acland G.M., Ray J.;
RT   "Cloning and expression of type II collagen mRNA: evaluation as a
RT   candidate for canine oculo-skeletal dysplasia.";
RL   Gene 255:307-316(2000).
DR   EMBL; AF023169; AAC62178.2; -.
DR   GO; GO:0005581; C:collagen; IEA.
DR   GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD000007; Clg_helix; 4.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
KW   Collagen.
SQ   SEQUENCE   1487 AA;  141875 MW;  25873EAC1E311DB8 CRC64;

Query Match      72.5%;  Score 2225;  DB 6;  Length 1487;
Best Local Similarity   72.6%;  Pred. No. 2.4e-125;
Matches  395;  Conservative  38;  Mismatches  111;  Indels   0;  Gaps   0;
```

```
Qy    1  GSEGPEGVRGEPGPPGPAGPAGAAGPADGAGPDGADGEPGAKGADGAGPGIAGAPGPFGARGPSGPE   60
              || | |  |||| ||  | |  | |  | | || | || || | | ||| |||||  || | |:
Db  384  GPEGAQGPRGEPGEPGTPGSPGAGASGNPGTDGIPGAKGSAGAPGIAGAPGFPGPRGPPGPGPGPQ  443

Qy   61  GPGGPPGKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGAGEEGKPGEPGARGEPGARGEPGAGPV  120
              | |||||   || | | | | | |||||   |||  ||| |||| |||||  || | ||| |||
Db  444  GATGPLGPKGQTGEPGIAGFKGEQGPKGEPGAPGPQGAPGPQGAPGPAGEEGKRGARGEPGGAGPV  503

Qy  121  GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAGAGLPGAKGLT  180
              | || || ||  | |  | |  |  |:|| | || | |  |  |   || | |  ||| | |||| :
Db  504  GPPGERGAPGNRGFPGQDGLAGPKGAPGERGPSGLAGPKGANGDGRPGEPGLPGARGLT  563

Qy  181  GSPGSPGDGKTGPPGPAGEDGRPGPPGPGKGEAGVMGFPGPKGAAGEPGKAGERGVP  240
              ||:||| |  |||| || |  |||| ||| |:|| |  |||||  |||||| | ||||
Db  564  GRPGDAGPQGKVGPSGAPGEDGRPGPGFPGPQGAPGPAGERGEQGRGAGPGEPGKAGEKGLP  623

Qy  241  GPPGAVGPAGKDGEAGAEGPPGPEAGPGAGKPGEE  300
              | ||  | |||| | ||| |||  | |  | ||| |:
Db  624  GAPGLRGLPGKDGBTGAAGPGPPGAGERGEQGAPGPG  683

Qy  301  GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPAGDGAPGADGDDGAKGDAGAPGAPGSE  360
              ||:| ||  |||  ||||  ||||| | |||| ||   | | |   |||  | ||||| ||||
Db  684  GVPGEAGAPGLVGPRGERGFPGERGSPGAQGLQGPRGGLPGTPGTDGPKGAASGPAGPPGAQ  743

Qy  361  GAPGLEGMPGERGAAGLPGPKGDGADGAGPCDRGEPGPPGPAGFAGPPGDAGEPGAKGDAGPP  420
              || | | ||||||  |||   | |  | |    || | ||  |  | |  | || | ||| |||
Db  744  GPPGLQGMPGERGAAGIAGPKGDRGDVGEKGPEGAPGKDGDGGRGLTGPIGPPGPAGPAGANGEK  803

Qy  421  GESGPSGPAGPTGARGAPGPCDRGEPGPPGPAGFAGPPGDAGEPGAKGDAGPP  480
              ||:||:||| | ||||| | ||  |||| || | | || |   |  ||  | | |
Db  804  GEVGPPGPAGTAGARGAPGERGETGPPGPAGFAGPPGADGQPGAKGEQGEAGQQKDGAGAP  863

Qy  481  GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGSGDAGPPGPPGPA  540
              || ||  |  | | | | || |  ||| || ||| | |  |||| || | |||||| | ||
Db  864  GPQGPSGAPGPQGPGTGPGVTGPKGARGAQGPPGATGFPGAAGRVGPPGSNGNPGPPGPPGPS  923

Qy  541  GKEG  544
              ||::|
Db  924  GKDG  927
```

```
RESULT 11
Q14046
ID   Q14046        PRELIMINARY;       PRT;   1160 AA.
AC   Q14046;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   COL2A1 protein precursor (Fragment).
GN   COL2A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cartilage;
RX   MEDLINE=90026318; PubMed=2803268;
RA   Baldwin C.T., Reginato A.M., Smith C., Jimenez S.A., Prockop D.J.;
RT   "Structure of cDNA clones coding for human type II procollagen. The
RT   alpha 1(II) chain is more similar to the alpha 1(I) chain than two
RT   other alpha chains of fibrillar collagens.";
RL   Biochem. J. 262:521-528(1989).
DR   EMBL; X16711; CAA34683.1; -.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD000007; Clg_helix; 4.
KW   Collagen; Signal.
FT   SIGNAL         1      23       POTENTIAL.
FT   CHAIN        113   >1160       COLLAGEN.
FT   NON_TER     1160    1160
SQ   SEQUENCE   1160 AA;  105630 MW;  A7F0523B56C8639 CRC64;
```

```
Query Match           72.4%;   Score 2224;  DB 4;   Length 1160;
Best Local Similarity 72.6%;   Pred. No. 2.2e-125;
Matches 395; Conservative 38; Mismatches 111; Indels 0; Gaps 0;

Qy     1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE  60
           ||||:||| |||||:|||| ||:|:|:|| ||| :| ||||||||||||||:|| ||:
Db   315 GPEGAQQPRGEPGTPGSPGPAGASGNPGTDGIPGAKGSAGAPGIAGAPGFPGPRGPPGPQ 374

Qy    61 GPGGPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVGPPGPGAGEGKPGARGEPGPTGLP  120
           ||||||:||:|||    |:  :| |||||||    |: ||:||:|||:||||| ||:
Db   375 GATGPLGPKGQTGEPGLAGFKGEQGPKGEPGPAGPQGPAGPAGEGKRGARGEPGGVGPI  434

Qy   121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
           |||:|| |||:|: |||   |:|||  ||||:||| :|| :||||:|||||||| ||||
Db   435 GPPGERGAPGNRGFPPGQDGLAGPKGAPGERGPSGLLAGPKGANGDPGRPGEPGLPGARGLT 494

Qy   181 GSPGSPGDGKTGPPGENGENGPGPGPGPGPARGEAGVMGFPGPKGAAGEPGKAGERGVP  240
           ||||:||||:||| |||:||:||| ||||||:||  |||||||||||:||||:|| ||:
Db   495 GRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGPGVMGFPGPKGANGEPGKAGEKGLP     554

Qy   241 GPPGAVGPAGKDGEAGAGEPGPGPAGPAGERGEGPAGSPGFEGLPGPAGPPGEAGKPGEE 300
           | ||  || ||||:|:|||||:||||||||||| |||||||:||||| |:||| |||::
Db   555 GAPGLRGLPGKDGETGAAGPPGPAGPAGERGESQGAPGSPGSGFQGLPGPGPGPGEGGKPGDQ 614

Qy   301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGDADGAPGDDGKAGDAGAPGAPGSE 360
           ||||: |||| |:||| ||||||| ||||||||||||||| ||| ||  ||:||||| ||:
Db   615 GVPGEAGAPGLVGPRGERGFPGERGSPGAQLQGPRGLPGTPGTDGPGKASGPAGPPGAQ   674

Qy   361 GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
           ||||| ||||||||| |||||| ||| || ||  |:|||:|||||||||||||||||:|:
Db   675 GPPGLQGMPGERGAAGIAGPKGDKGRDGDVGEKGPEGAPGKDGKGRGLTGPIGPPGPAGANGEK 734

Qy   421 GESGPSGPAGPTGARGARGPGDREGPGPPGPAGPAGPPGPGADGEPGAKGEPGDAGAKGDAGPP 480
           ||:|| ||||:||| || | || |||||||||| | ||:||||:||| ||  |||||||||||
Db   735 GBVGPPGPAGSAGARGAPGERGETGPPGPAGFAGPPPGADGQPGAKGBQGEAGQKGDAGAP     794

Qy   481 GPAGPAGPPGPIGDVGPAPGAKGARGSAGPPGPATGFPGAAGRVGPPGPSGDAGPPGPGPA 540
           | ||||||||| || || ||||||| |||||||:||||||||||||||||:|| ||||||:
Db   795 GPQGPSGAPGPQGPQGPGTVGTGPKGARGAQGPPGAATGFPGAAGRVGPPGPSGNGNPGCPPGPPGPS 854

Qy   541 GKEG 544
           ||:|
Db   855 GKDG 858


RESULT 12
Q14047
ID   Q14047      PRELIMINARY;      PRT;   1487 AA.
AC   Q14047;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Alpha-1 type II collagen.
GN   COL2A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=85190534; PubMed=3857598;
RA   Cheah K.S., Stoker N.G., Griffin J.R., Grosveld F.G., Solomon E.;
RT   "Identification and characterization of the human type II collagen
RT   gene (COL2A1).";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:2555-2559 (1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=90026318; PubMed=2803268;
RA   Baldwin C.T., Reginato A.M., Smith C., Jimenez S.A., Prockop D.J.;
RT   "Structure of cDNA clones coding for human type II procollagen. The
RT   alpha 1(II) chain is more similar to the alpha 1(I) chain than two
RT   other alpha chains of fibrillar collagens.";
RL   Biochem. J. 262:521-528(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=89325561; PubMed=2753125;
RA   Vikkula M., Peltonen L.;
RT   "Structural analyses of the polymorphic area in type II collagen
RT   gene.";
RL   FEBS Lett. 250:171-174(1989).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=91184811; PubMed=2081599;
RA   Ryan M.C., Sieraski M., Sandell L.J.;
RT   "The human type II procollagen gene: identification of an additional
RT   protein-coding domain and location of potential regulatory sequences
RT   in the promoter and first intron.";
RL   Genomics 8:41-48(1990).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=91153296; PubMed=1999183;
RA   Huang M.C., Seyer J.M., Thompson J.P., Spinella D.G., Cheah K.S.,
RA   Kang A.H.;
RT   "Genomic organization of the human procollagen alpha 1(II) collagen
RT   gene.";
RL   Eur. J. Biochem. 195:593-600(1991).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=92344585;
RA   Vikkula M., Metsaranta M., Syvanen A.C., Ala-Kokko L., Vuorio E.,
RA   Peltonen L.;
RT   "Structural analysis of the regulatory elements of the type-II
RT   procollagen gene. Conservation of promoter and first intron sequences
RT   between human and mouse.";
RL   Biochem. J. 285:0-0(0).
RN   [7]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=97104294; PubMed=8948452;
RA   Ala-Kokko L., Kvist A.P., Metsaranta M., Kivirikko K.I.,
RA   de Crombrugghe B., Prockop D.J., Vuorio E.;
RT   "Conservation of the sizes of 53 introns and over 100 intronic
RT   sequences for the binding of common transcription factors in the human
RT   and mouse genes for type II procollagen (COL2A1).";
RL   Biochem. J. 308:0-0(0).
DR   EMBL; L10347; AAC41772.1; -.
DR   GO; GO:0005581; C:collagen; IEA.
DR   GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD000007; Clg_helix; 4.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS01384; VWFC_2; 1.
KW   Collagen.
SQ   SEQUENCE 1487 AA; 141772 MW; 0B7E79D46BDAFA97 CRC64;

Query Match           72.4%;   Score 2224;   DB 4;   Length 1487;
Best Local Similarity 72.6%;   Pred. No. 2.8e-125;
Matches 395; Conservative 38; Mismatches 111; Indels 0; Gaps 0;

Qy     1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE 60
```

```
KW   Collagen.
SQ   SEQUENCE   1418 AA;   134343 MW;   115FCD19EB8696A3 CRC64;

Query Match            72.3%;   Score 2220;   DB 6;   Length 1418;
Best Local Similarity  72.6%;   Pred. No. 4.6e-125;
Matches 395;   Conservative 37;   Mismatches 112;   Indels 0;   Gaps 0;

Qy      1 GSEGPEGVRGEPGPGPPGAGAAGPAGPGADGEPGDEAKGDGADGAPGIAGAGPFGARGPSGPE   60
          | | | : || |||| :: |:  | |||||||| |||||| : :: |||| ||||||||||:
Db    315 GPEGAQGPRGEPGTPGSPGPAGASGNPGTDGIPGAKGSAGAPGIAGAGPFGGPRGRGPPGPQ  374

Qy     61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGGARGEPGGPTGLP            120
          |||| ||| || : | || |: ||| :: :: :||||| :|||||||| :: |||
Db    375 GATGPLGPKGQTGEPGIAGFKGBQGPKGEPGPAGPQAPGPAGEGEGKRGARGEPGGVGVPI  434

Qy    121 GPPGERGGRGSPGPADGVAGPKGPGAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT     180
          |||||| :: | ||||| :: | | ||:: : ||| || : || :: |||| |||||
Db    435 GPPGERGAPGNRGFPGQDGLAGPKGAPGERGPSGLLAGPKGANGDPGRPGEPGLPGARGLT  494

Qy    181 GSPGSPGFGPGEDGRPGPPGPAGEDGRPGPGPGPGARGEAGVMGFPGPKGAAGEPGKAGRGVP  240
          |||| || : | || ::: || |||| |: : | | | | | |||||| : |||||||
Db    495 GRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGPGPGARGPSGLLAGPKGEPDGKAGEKGLP  554

Qy    241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEEGPAGSPGPEGLPGPAGPPGEAGKPGEE    300
          ||||| | || :: : | | |||| || | :||||| | :: |:: ||||| : ::
Db    555 GAPGLRGLPGKDGETGAAGPPGPAGEDGRPGPGPGDGAKGDAGAPGAPGSE  614

Qy    301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGPPGDAKGEPGAKGPDAGAPGAPGSE     360
          |||||| :: |||||| | | || |:: :: |||||||||| || |||| |
Db    615 GVPGEAGAPGLVGPRGERGFPGERGSPGAQGLQGARGLPGTPGTDGFKGASGPAGSPGAQ  674

Qy    361 GAPGLEGMPGERGAAGLPGPKGDRDGAPGSPGKDGVRGLTGPIGPPGPAGAPGDK         420
          |||||: || || :: | | ||||| :|||| || :| |: ||||||
Db    675 GPGLQGMPGERGAAGIAGPKGDRDGDVGEKGPEGAPGKDGKGRGLTGPIGPPGPAGANGEK  734

Qy    421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGERGETGPPGAPGAPAGPPGDADGSPGKDGVGPLTGPIGPPGPAGAPGDK  480
          |||| ||| ||||| || :|| |||||||||||| :: | |||||| : : || ||
Db    735 GEVGPPGPSGAGARGAPGERGETGPPGAPGAPGDGQPGAKGEQGEAQGKGDAGAP  794

Qy    481 GPAGPAGPPGPIGDVGAPAGKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPPGPA       540
          :: :: |||| :|| ||||||| | || |||||| : ||: ||||| ||||
Db    795 GPQGPSGAPGNGPQGPTGVTGPKGARGAQGPPGATGFPGAAGRVGPPGSNGNPGPPGPS  854

Qy    541 GKEG   544
          |||:
Db    855 GKDG   858
```

RESULT 14
Q7T2Z7
ID   Q7T2Z7;   PRELIMINARY;   PRT;   1269 AA.
AC   Q7T2Z7;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Alpha 1 type II procollagen (Fragment).
GN   COL2A1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Caixia X., Yongzhi X., Siqi G., Yuying S.;
RT   "Gallus gallus alpha 1 type IIA collagen precursor (COL2A1).";
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF452711; AAO33039.2; -.
KW   Collagen.
FT   NON_TER   1   1
SQ   SEQUENCE   1269 AA;   120116 MW;   446EF91BEBBBBF6A CRC64;

Query Match   71.5%;   Score 2195;   DB 13;   Length 1269;

(Second alignment column, left side of page)

```
Db    384 GPEGAQGPRGEPGTPGSPGPAGASGNPGTDGIPGAKGSAGAPGIAGAGPFGGPRGRGPPGPQ  443
          | |: | |||| : |: || ||| :: :::: |||| ||||||||||
Qy     61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGGARGEPGGPTGLP            120
          |||| ||| || | | || |: ||: :: : ||||| ||||||| : :: |||
Db    444 GATGPLGPKGQTGEPGIAGFKGBQGPKGEPGPAGPQAPGPAGEGEGKRGARGEPGGVGVPI  503

Qy    121 GPPGERGGRGSPGPADGVAGPKGPGAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT     180
          |||||| :: | ||||| :: | | ||:: : ||| || :: || :: |||| |||||
Db    504 GPPGERGAPGNRGFPGQDGLAGPKGAPGERGPSGLLAGPKGANGDPGRPGEPGLPGARGLT  563

Qy    181 GSPGSPGFGPGEDGRPGPPGPAGEDGRPGPGPGPGARGEAGVMGFPGPKGAAGEPGKAGRGVP  240
          |||| || : | || ::: || |||| |: : | | | | | |||||| : |||||||
Db    564 GRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGARGQPGPGPGARGPSGLLAGPKGEPDGKAGEKGLP  623

Qy    241 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGPPGDAKGEPGAKGPDAGAPGAPGSE    300
          |||||| :: |||||| | | || |:: :: |||||||||| || |||| |
Db    624 GAPGLRGLPGKDGETGAAGPPGPAGEDGRPGPGPGDGAKGDAGAPGAPGSE  683

Qy    301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGPPGDAKGEPGAKGPDAGAPGAPGSE     360
          |||||| :: |||||| | | || |:: :: |||||||||| || |||| |
Db    684 GVPGEAGAPGLVGPRGERGFPGERGSPGAQGLQGARGLPGTPGTDGFKGASGPAGSPGAQ  743

Qy    361 GAPGLEGMPGERGAAGLPGPKGDRDGAPGSPGKDGVRGLTGPIGPPGPAGAPGDK         420
          |||||: || || :: | | ||||| :|||| || :| |: ||||||
Db    744 GPGLQGMPGERGAAGIAGPKGDRDGDVGEKGPEGAPGKDGKGRGLTGPIGPPGPAGANGEK  803

Qy    421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGERGETGPPGAPGAPAGPPGDADGSPGKDGAGPP  480
          |||| ||| ||||| || :|| |||||||||||| :: | |||||| : : || ||
Db    804 GEVGPPGPSGAGARGAPGERGETGPPGAPGAPGDGQPGAKGEQGEAQGKGDAGAP  863

Qy    481 GPAGPAGPPGPIGDVGAPAGKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPPGPA       540
          :: :: |||| :|| ||||||| | || |||||| : ||: ||||| ||||
Db    864 GPQGPSGAPGNGPQGPTGVTGPKGARGAQGPPGATGFPGAAGRVGPPGSNGNPGPPGPS  923

Qy    541 GKEG   544
          |||:
Db    924 GKDG   927
```

RESULT 13
Q28396
ID   Q28396;   PRELIMINARY;   PRT;   1418 AA.
AC   Q28396;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Type II collagen.
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Richardson D.W., Dodge G.R.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 18-68 FROM N.A.
RA   MacLeod J.N., Fubini S.L., Gu D.N., Tetreault J.W., Todhunter R.J.;
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U62528; AAB05773.1; -.
DR   EMBL; AF040638; AAB96768.1; -.
DR   PIR; T45467; T45467.
DR   GO; GO:0005581; C:collagen; IEA.
DR   GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD000007; Clg_helix; 4.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.

Best Local Similarity 71.7%; Pred. No. 1.3e-123; Indels 0; Gaps 0;
Matches 390; Conservative 37; Mismatches 117;

```
Qy    1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE   60
          | ::| : ||  :|||||||  ||| |||  |||||||||| |||||||||||| ||| |
Db  166 GPEGAQGPRGEGSGTPGSPGPAGAPGNPGTDGIPGAKGSAGAPGIAGAPGFPGPRGPPGPQ  225

Qy   61 GPGGPPGKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEBGKFGARGEPGPTGLP   120
          |  ||||| |||||| || |||  || |:|| | |  ||||||| |:|| || |||||||
Db  226 GATGPLGPKGQTGEPGIAGFKGEQGPKGETGPAGPQGAPGPAGEGKRGARGEPGAAGPV    285

Qy  121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT  180
          |||||||  | :|| ||||   |||| ||||| :  ||||  | |||| ||| |||:|||
Db  286 GPPGERGAPGNRGFPGQDGLAGPKGAPGERGPSLAGPKGANGDPGRPGEPGLPGARGLT   345

Qy  181 GSPGSPGPDGKTGPPGPAGEDGRPGRPGP2GPGPQGPARGQPGVMGFPGPKGANGEPGKAGEKGLP  240
          | ||||||  | :||| ||||:|||| |  |||  ||  || |:||||||| |||| ||:|| ||
Db  346 GRPGDAGPQGKVGPTGAPGEDGRPGRPGP2GPGPQGPARGQPGVMGFPGPKGANGEPGKAGEKGLP  405

Qy  241 GPPGAVGPAGKDGEAGAEGPPGPAGERGEEGPAGSPGFEGLPGAGPPGEAGKPGEE     300
          |||||||   || || ||||||| ||| || | | ||  || |||| |||||| ||
Db  406 GAPGLRGLPGKDGETGAAGPPGPGPAGPVGERGEQGAPGPVGERGEQGAPGSGFQGLPGPPGPPGESGKPGDQ  465

Qy  301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGPGADGAPGDDGAKGDAGAPGSE    360
          | |||||| || || | ||||| |  || || || || | || | || ||||| | |:
Db  466 GVPGEAGAPGLVGPRGERGFPGERGSPGAQLQGPRGLPGTPGTDGPKGATGPAGMGPGDK  525

Qy  361 GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  420
          ||| ||| |||||| | ||||| ||||| | || |||| ||| |||||||||||||||||
Db  526 GPPGLQGMPGERGAAGIAGPKGDRDGVEKGPEGAPGPAGPPGADGQPGADGQPGAKGDQGP  585

Qy  421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPPGPGADGEPGDAGAKGEPGDAGAKDGADGPP  480
          || | || || || || | || |||||| ||| | | ||||| || | | || | || |||
Db  586 GESGPSGPAGPTGARGAPGDRGEPGPPGPGADGCPGADGQPGADGQPGADGQPGAKGDQGPGPGDAGAGP  645

Qy  481 GPAGPAGPPGIGDVGAPGAKGARGSAGPPGATGPPGAAGRVGPPGATGPPGAAGRVGPPGSA  540
          || ||||| | | |||| || || | || || || |||| |||| || ||||| | ||| |
Db  646 GPQGPSGAPGPQGPGPTGVTGPKGARGAQGPPGATGPPGNGNPGPGPPGSA            705

Qy  541 GKEG  544
          ||:|
Db  706 GKDG  709
```

RESULT 15
Q63123
ID   Q63123   PRELIMINARY;   PRT;   1419 AA.
AC   Q63123; Q63565;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Collagen alpha 1 type II (T1 mRNA).
GN   T1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Callus;
RA   Urabe K., Sarkar G., Bolander M.E.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1372-1419 FROM N.A.
RA   Wurtz T., Brandsten C., Lundmark C., Christersson C.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 370-422 FROM N.A.
RC   STRAIN=DA; TISSUE=Cartilage;
RX   MEDLINE=94321934; PubMed=8046350;
RA   Michaelson E., Malmstrom V., Reis S., Engstrom A., Burkhardt H.,
RA   Holmdahl R.;
RT   "T cell recognition of carbohydrates on type II collagen.";

RL   J. Exp. Med. 180:745-749(1994).
DR   EMBL; L48440; AAA79780.1; -.
DR   EMBL; AJ224879; CAA12179.1; -.
DR   EMBL; X79816; CAA56213.1; -.
DR   PIR; I60384; I60384.
DR   GO; GO:0005581; C:collagen; IEA.
DR   GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD000007; Clg_helix; 4.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
KW   Collagen.
SQ   SEQUENCE 1419 AA; 134570 MW; B7C63B77819CB50B CRC64;

Query Match   71.4%; Score 2192; DB 11; Length 1419;
Best Local Similarity   71.7%; Pred. No. 2.2e-123;
Matches 390; Conservative 40; Mismatches 114; Indels 0; Gaps 0;

```
Qy    1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE   60
          | ::| : ||  ||| ||  ||| |||  ||||||||| |||||||||||| ||| :
Db  316 GPEGAQGSRGEPGNPGSPGPAGASGNPGTDGIPGAKGSAGAPGIAGAPGFPGPRGPPGPQ  375

Qy   61 GPGGPPGKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEBGKFGARGEPGPTGLP   120
          |  | | |  |||||| || |||  || | || | |  ||||||| |:|| || |||||
Db  376 GATGPLGPKGQTGEPGIAGFKGEQGPKGETGPAGPQGAPGPQGAPGPAGEEGKRGARGEPGGAGPI  435

Qy  121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT  180
          |||||||  | :|| ||||   |||| ||||  |  ||||  |  ||||| ||| |||:|
Db  436 GPPGERGAPGNRGFPGQDGLAGPKGAPGERGPSLAGPKGANGDPGRPGEPGLPGARGLT   495

Qy  181 GSPGSPGPDGKTGPPGPAGEDGRPGRPGP2GPGPQGPARGQPGVMGFPGPKGANGEPGKAGERGVP  240
          | |||||| | :||| ||||:|||| |  |||  ||  || |:||||||| |||| |  |:
Db  496 GRPGDAGPQGKVGPSGAPGEDGRPGPPGPQGQGARGQPGVMGFPGPKGANGBPGANGEKGLA  555

Qy  241 GPPGAVGPAGKDGEAGAEGPPGPAGERGEEGPAGSPGFEGLPGAGPPGEAGKPGEE     300
          |||||||   || || ||||||| ||| || | | ||  ||| |||| |||||| ||
Db  556 GAPGLRGLPGKDGETGAAGPPGSPGPAGERGEQGAPGDSGFQGLPGPDPGEGGQKGDQ    615

Qy  301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGPGADGAPGDDGAKGDAGAPGSE    360
          | |||||| || || | ||||| |  || || || || | || | || |||||| | |:
Db  616 GIPGEAGAPGLVGPRGERGFPGERGSPGAQLQGPRGLPGTPGTDGPKGAAGPDGPPGAQ   675

Qy  361 GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  420
          ||| ||| |||||| | ||||| ||||| | || |||| ||| ||||||||||| ||||
Db  676 GPPGLQGMPGERGAAGIAGPKGDRGDVGEKGPEGAPGFAGPDGEPGAKGEPGDAGAKGDAGPP  735

Qy  421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPPGPGADGEPGDAGAKGEPGDAGAKDGADGP  480
          || | || || || || | || |||||| | || | |||||||  |  | | || | |||
Db  736 GEVGPPGPSGSSTGARGAPGERGETGPPGPAGPAGPGADGQPGADGQPGEAGQKGDAGAP   795

Qy  481 GPAGPAGPPGIGDVGAPGAKGARGSAGPPGATGPPGAAGRVGPPGSGDAGPPGSPGPGPPA  540
          || ||||| | | |||| || || | |||| ||| | || |||| | || ||| | ||| |
Db  796 GPQGPSGAPGPQGPGPTGVTGPKGARGAQGPPGATGPPGAAGRVGPGSNGNPGPAGPTGPA  855

Qy  541 GKEG  544
          ||:|
Db  856 GKDG  859
```

Search completed: May 7, 2004, 16:29:49
Job time : 48 secs

```
                    GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    May 7, 2004, 16:22:49 ; Search time 59 Seconds
                              (without alignments)
                           2605.184 Million cell updates/sec

Title:         US-10-658-989A-4
Perfect score: 3070
Sequence:      1 GSEGPEGVRGEPGPPGPGPAGA........PGPSGDAGPPGPGPAGKEG 544

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:    1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_29Jan04:*
              1: geneseqp1980s:*
              2: geneseqp1990s:*
              3: geneseqp2000s:*
              4: geneseqp2001s:*
              5: geneseqp2002s:*
              6: geneseqp2003as:*
              7: geneseqp2003bs:*
              8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2966	96.6	1057	3	AAY84541	Aay84541 Amino aci
2	2966	96.6	1057	3	AAY84544	Aay84544 A human c
3	2966	96.6	1058	3	AAY84403	Aay84403 Amino aci
4	2966	96.6	1107	2	AAR89472	Aar89472 Collagen/
5	2966	96.6	1107	3	AAY84540	Aay84540 Amino aci
6	2966	96.6	1161	7	ADE87050	Ade87050 Human pan
7	2966	96.6	1169	2	AAR89469	Aar89469 Collagen/
8	2966	96.6	1169	2	AAY84537	Aay84537 Amino aci
9	2966	96.6	1171	2	AAR89470	Aar89470 Collagen/
10	2966	96.6	1171	2	AAY84538	Aay84538 A chimeri
11	2966	96.6	1211	7	ADE87057	Ade87057 Human pan
12	2966	96.6	1226	7	ADE87062	Ade87062 Human pan
13	2966	96.6	1388	3	AAR89471	Aar89471 Collagen/
14	2966	96.6	1411	3	AAY56800	Aay56800 Human pre
15	2966	96.6	1461	5	ABG93947	Abg93947 Human pol
16	2966	96.6	1464	5	AAW68485	Aaw68485 Human rec
17	2966	96.6	1464	4	AAB82454	Aab82454 Human pro
18	2966	96.6	1464	4	AAU14136	Aau14136 Human nov
19	2966	96.6	1464	5	ABB90764	Abb90764 Human Tum
20	2966	96.6	1464	5	ABP68610	Abp68610 Human pan
21	2966	96.6	1464	6	ABU54471	Abu54471 Human tum
22	2966	96.6	1464	6	ABR47417	Abr47417 Breast ca
23	2966	96.6	1464	6	ABR92064	Abr92064 Human cer
24	2966	96.6	1464	6	ADD14142	Add14142 Human src
25	2966	96.6	1464	7	ADD45059	Add45059 Human Pro
26	2966	96.6	1464	7	ADD45055	Add45055 Human Pro
27	2966	96.6	1464	7	ADD45051	Add45051 Human Pro
28	2958	96.6	1464	3	ADE87048	Ade87048 Human pan
29	2958	96.4	1388	3	AAY84539	Aay84539 Amino aci
30	2936	95.6	1463	4	AAE02532	Aae02532 Bovine al
31	2920	95.1	1536	7	ADE87051	Ade87051 Human pan
32	2872	93.6	1453	7	ADD45053	Add45053 Rat Prote
33	2872	93.6	1453	7	ADD45057	Add45057 Rat Prote
34	2872	93.6	1453	7	ADD48341	Add48341 Rat Prote
35	2872	93.6	1453	7	ADD45049	Add45049 Rat Prote
36	2872	93.6	1453	7	ADD48337	Add48337 Rat Prote
37	2872	93.6	1453	7	ADD48345	Add48345 Rat Prote
38	2857	93.1	822	2	AAY06240	Aay06240 Mouse rec
39	2857	93.1	1453	5	ABG93948	Abg93948 Mouse pol
40	2816.5	91.7	1449	4	AAE02535	Aae02535 Porcine a
41	2791.5	90.9	1341	2	AAR71701	Aar71701 Collagen
42	2791.5	90.9	1341	3	AAY96122	Aay96122 Collagen
43	2791.5	90.9	1341	5	AAE16475	Aae16475 Human col
44	2791.5	90.9	1341	5	ABB80733	Abb80733 Collagen
45	2791.5	90.9	1341	5	ABB09625	Abb09625 Amino aci

ALIGNMENTS

```
RESULT 1
AAY84541
ID   AAY84541 standard; protein; 1057 AA.
XX
AC   AAY84541;
XX
DT   25-JUL-2000  (first entry)
XX
DE   Amino acid sequence of a human collagen 1 (alpha1) protein.
XX
KW   Extracellular matrix protein; self aggregation; hydroxylated proline;
KW   trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
KW   collagen; fibrinogen; fibronectin; post translational hydroxylation.
XX
OS   Homo sapiens.
XX
PN   EP992586-A2.
XX
PD   12-APR-2000.
XX
PF   07-OCT-1999;   99EP-00119184.
XX
PR   09-OCT-1998;   98US-00169768.
XX
PA   (USSU ) US SURGICAL CORP.
XX
PI   Gruskin EA, Buechter DD, Zhang G, Connolly K;
XX
DR   WPI; 2000-259138/23.
DR   N-PSDB; AAA12502.
XX
PT   Production of extracellular matrix proteins containing 4-trans-
PT   hydroxyproline results in native self aggregating proteins, useful on
PT   medical implants.
XX
PS   Disclosure; Fig 27A-E; 260pp; English.
XX
CC   The specification describes a method for producing an extracellular
CC   matrix protein or its fragment. The extracellular matrix protein is
CC   capable of self aggregating in a cell which does not ordinarily
CC   hydroxylated prolines. The method comprises optimising a nucleic acid
CC   sequence for expression in the cell by substitution of codons preferred
CC   by that cell for naturally occurring codons not preferred by the cell;
CC   incorporating the nucleic acid sequence into the cell; and contacting the
CC   cell with a hypertonic growth medium containing at least one amino acid,
CC   selected from the group consisting of trans-4-hydroxyproline and 3-
CC   hydroxyproline to allow at least one of the amino acids to be assimilated
CC   into the cell and incorporated into the extracellular matrix protein. The
```

Right column

```
PN   EP992586-A2.
XX
PD   12-APR-2000.
XX
PF   07-OCT-1999;   99EP-00119184.
XX
PR   09-OCT-1998;   98US-00169768.
XX
PA   (USSU ) US SURGICAL CORP.
XX
PI   Gruskin EA, Buechter DD, Zhang G, Connolly K;
XX
DR   WPI; 2000-259138/23.
DR   N-PSDB; AAA12503.
XX
PT   Production of extracellular matrix proteins containing 4-trans-
PT   hydroxyproline results in native self aggregating proteins, useful on
PT   medical implants.
XX
PS   Example 10; Fig 39A-E; 260pp; English.
XX
CC   The specification describes a method for producing an extracellular
CC   matrix protein or its fragment. The extracellular matrix protein is
CC   capable of self aggregating in a cell which does not ordinarily
CC   hydroxylated prolines. The method comprises optimising a nucleic acid
CC   sequence for expression in the cell by substitution of codons preferred
CC   by that cell for naturally occurring codons not preferred by the cell;
CC   incorporating the nucleic acid sequence into the cell; and contacting the
CC   cell with a hypertonic growth medium containing at least one amino acid,
CC   selected from the group consisting of trans-4-hydroxyproline and 3-
CC   hydroxyproline to allow at least one of the amino acids to be assimilated
CC   into the cell and incorporated into the extracellular matrix protein. The
CC   method may be used to make host cells assimilate and incorporate trans-4-
CC   hydroxyproline into proteins. This is especially useful in the
CC   recombinant production of proteins such as collagen, fibrinogen and
CC   fibronectin whose ability to self aggregate and produce functional
CC   proteins depends on the post translational hydroxylation of proline. The
CC   method is also useful in studying the structure and function of
CC   polypeptides which do not normally contain trans-4-hydroxyproline. The
CC   present sequence represents human collagen 1 (alpha1) helical region,
CC   which may be produced using the method of the invention
XX
SQ   Sequence 1057 AA;

Query Match             96.6%;  Score 2966;  DB 3;  Length 1057;
Best Local Similarity   95.6%;  Pred. No. 2.7e-169;
Matches   520;  Conservative  21;  Mismatches  3;  Indels  0;  Gaps  0;

Qy    1 GSEGPEGVRGEPGPPGPAGAAGPADGPGAKGADGAPGIAGAPGFPGARGPSGPE       60
          ||||| :|||||||||||||||||| :||||||||| :|||||||||||||||||:
Db  201 GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ 260

Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEBGKPGARGEPGPTGLP 120
          |||||||||: |||||||||||||||||||||||: |||||||||| :|: |||||||||
Db  261 GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEEGKRGARGEPGPTGLP 320

Qy  121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  321 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 380

Qy  181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPARGEAGVMGFPGPKGAAGEPGKAGERGVP   240
          |||||||||||||||||||: ||||||||||| :||||||||||||||||||||||||
Db  381 GSPGSPGPDGKTGPPGPAGQDGRPGPPGPARGQAGVMGFPGPKGAAGEPGKAGERGVP   440

Qy  241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEBGPAGSPGFPGADGPPGPAGPPGEAGKPGEE 300
          |||||||||||||||| :||||||||||||  : ||||||| : |||||||||||||||| :
Db  441 GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEGQGPAGSPGFQGLPGPAGPPGEAGKPGEQ   500

Qy  301 GVPGDLGAPGSGARGEPGFPGERGVEGPPGPAGEDGRPGPPGPAGPPGPPGPPGPPGPA   360
          |||||||||||||||: |||||||: |||||||  : |||  : ||  : |  :  ||
Db  501 GVPGDLGAPGSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDAGAPGAPGAPGSQ     560

Qy  361 GAPGLEMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
          ||||| :||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  561 GAPGLQMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 620
```

Left column

```
CC   method may be used to make host cells assimilate and incorporate trans-4-
CC   hydroxyproline into proteins. This is especially useful in the
CC   recombinant production of proteins such as collagen, fibrinogen and
CC   fibronectin whose ability to self aggregate and produce functional
CC   proteins depends on the post translational hydroxylation of proline. The
CC   method is also useful in studying the structure and function of
CC   polypeptides which do not normally contain trans-4-hydroxyproline. The
CC   present sequence represents a human collagen 1 (alpha1) protein, which
CC   may be produced using the method of the invention
XX
SQ   Sequence 1057 AA;

Query Match             96.6%;  Score 2966;  DB 3;  Length 1057;
Best Local Similarity   95.6%;  Pred. No. 2.7e-169;
Matches   520;  Conservative  21;  Mismatches  3;  Indels  0;  Gaps  0;

Qy    1 GSEGPEGVRGEPGPPGPAGAAGPADGPGAKGADGAPGIAGAPGFPGARGPSGPE       60
          ||||| :|||||||||||||||||| :||||||||| :|||||||||||||||||:
Db  201 GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ 260

Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEBGKPGARGEPGPTGLP 120
          |||||||||: |||||||||||||||||||||||: |||||||||| :|: |||||||||
Db  261 GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEEGKRGARGEPGPTGLP 320

Qy  121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  321 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 380

Qy  181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPARGEAGVMGFPGPKGAAGEPGKAGERGVP   240
          |||||||||||||||||||: ||||||||||| :||||||||||||||||||||||||
Db  381 GSPGSPGPDGKTGPPGPAGQDGRPGPPGPARGQAGVMGFPGPKGAAGEPGKAGERGVP   440

Qy  241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEBGPAGSPGFPGADGPPGPAGPPGEAGKPGEE 300
          |||||||||||||||| :||||||||||||  : ||||||| : |||||||||||||||| :
Db  441 GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEGQGPAGSPGFQGLPGPAGPPGEAGKPGEQ   500

Qy  301 GVPGDLGAPGSGARGEPGFPGERGVEGPPGPAGEDGRPGPPGPAGPPGPPGPPGPPGPA   360
          |||||||||||||||: |||||||: |||||||  : |||  : ||  : |  :  ||
Db  501 GVPGDLGAPGSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDAGAPGAPGAPGSQ     560

Qy  361 GAPGLEMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
          ||||| :||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  561 GAPGLQMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 620

Qy  421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGEPGADGAKGDAGPP           480
          |||||||||||: |||||||||||||||||||||||||||||||||||
Db  621 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGEPGADGQPGAKGDAGPP         680

Qy  481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPPGPPGPA 540
          |||||||||| :|||||||||||||||||||: ||||||||||||||||||||||||||
Db  681 GPAGPAGPPGPIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPA 740

Qy  541 GKEG 544
          ||||
Db  741 GKEG 744

RESULT 2
AAY84544
ID   AAY84544 standard; protein; 1057 AA.
XX
AC   AAY84544;
XX
DT   25-JUL-2000 (first entry)
XX
DE   A human collagen 1 (alpha1) protein helical region.
XX
KW   Extracellular matrix protein; self aggregation; hydroxylated proline;
KW   trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
KW   collagen; fibrinogen; fibronectin; post translational hydroxylation.
XX
OS   Homo sapiens.
XX
```

```
Matches  520;  Conservative  21;  Mismatches  3;  Indels  0;  Gaps  0;

Qy    1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPAGAKGADGAPGIAGAPGFPGARGPSGPE 60
        ||||:|||||||||||||||||||||:||||||:||||:||||||:|||||||:|
Db  202 GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPFPGARGPSGPQ 261

Qy   61 GPGGPPGKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEEGKPGARGEPGPTGLP 120
        |||||||||:|||||||||||||||||||||||:||||||||||:||||||||||:|||
Db  262 GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEEGKRGARGEPGPTCLP 321

Qy  121 GPGGERGPSGRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
        |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db  322 GPGGERGPSGRGFFGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 381

Qy  181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGARGBAGVMGFPGPKGAAGEPGKAGERGVP 240
        ||||||||||||||||||||:|||||||||||||:|||||||||||||||||||||||||
Db  382 GSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGVP 441

Qy  241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGEREBGPAGSPGFEGLPGAGPGEAGKPGEE 300
        ||||||||||||||||:||||||||||||:||:|||||||:|||||||||||||||:
Db  442 GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGFQGLPGAGPGEAGKPGEQ 501

Qy  301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGADGAPGDDGAKGDAGAPGAPGSE 360
        ||||||||||||||||||||||||||:|||||||||:||||||:|||||||||||||||:
Db  502 GVPGDLGAPGPSGARGEPGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQ 561

Qy  361 GAPGLEMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGAGAPGDK 420
        |||||:|||||||||||||||||||||||||||||:||||||||||||||||||:|||
Db  562 GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGDGVRGLTGPIGPPGAGADGDK 621

Qy  421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGDGEPGDAGAKGDAGPP 480
        |||||||||||||||||||||||||||||||||||:|||||||||||||
Db  622 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGDGQPGDAGAKGDAGPP 681

Qy  481 GPAGPAGPGPGTGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPGPPGPA 540
        ||||||||:|||:|||||||||||||||||||||||||||||||||||||:|||||||
Db  682 GPAGPAGPPGPIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPA 741

Qy  541 GKEG 544
        ||||
Db  742 GKEG 745
```

RESULT 4

AAR89472
ID   AAR89472 standard; protein; 1107 AA.
XX
AC   AAR89472;
XX
DT   01-OCT-1996   (first entry)
XX
DE   Collagen/decorin(aa46-93) fusion protein.
XX
KW   Transforming growth factor; TGF-beta-1; collagen IA; osteogenesis;
KW   bone formation; tissue repair; fusion protein.
XX
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Domain   1..1057
FT   /label= Collagen-IA
FT   /note= "Collagen-IA"
FT   Misc-difference 887
FT   /note= "unidentified amino acid"
FT   Misc-difference 890
FT   /note= "unidentified amino acid"
FT   Peptide   1058..1059
FT   /label= Linker_peptide
FT   Domain   1060..1107
FT   /label= Decorin
FT   /note= "amino acids P46 to G93 of mature decorin"
XX
PN   CA2151547-A.

```
Db  561 GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 620
        ||||:|||||||||||||||||||||||||||||||:|||||||||||||||||:||||
Qy  421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGDGEPGDAGAKGDAGPP 480
        ||||||||||||||||||||||||||||||||||||:|||||||||||||
Db  621 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGDGQPGDAGAKGDAGPP 680

Qy  481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPGPPGPA 540
        ||||||||||||:|||||||||||||||||||||||||||||||||||:||||||||
Db  681 GPAGPAGPPGPIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPA 740

Qy  541 GKEG 544
        ||||
Db  741 GKEG 744
```

RESULT 3

AAY84403
ID   AAY84403 standard; protein; 1058 AA.
XX
AC   AAY84403;
XX
DT   12-JUL-2000   (first entry)
XX
DE   Amino acid sequence of human type 1 (alpha1) collagen polypeptide.
XX
KW   Alpha1 collagen; 3,4-dehydro-L-proline; epoxidation; 3,4-epoxyproline;
KW   collagen; mussel adhesive protein; bioadhesive.
XX
OS   Homo sapiens.
XX
XX
PN   WO200014201-A1.
XX
PD   16-MAR-2000.
XX
PF   07-SEP-1999;   99WO-US020462.
XX
PR   09-SEP-1998;   98US-0099652P.
XX
PA   (USSU)   US SURGICAL CORP.
PA   (PAOL/)   PAOLELLA D N.
PA   (GRUS/)   GRUSKIN E A.
PA   (BUEC/)   BUECHTER D D.
XX
PI   Paolella DN, Gruskin EA, Buechter DD;
XX
DR   WPI; 2000-271051/23.
DR   N-PSDB; AAZ99843.
XX
PT   Incorporating non-natural amino acid into polypeptide, useful e.g. for
PT   production of bioadhesives, by epoxidation or substitution of
PT   dehydroproline residues.
XX
PS   Disclosure; Fig 6; 66pp; English.
XX
CC   The present sequence represents a human type 1 (alpha1) collagen protein.
CC   Peptides derived from the protien were used to demonstrate incorporation
CC   of 3,4-dehydro-L-proline into the peptide, using the method of the
CC   invention. The specification describes a method for the incorporation of
CC   non-natural amino acid into a polypeptide. The method comprises reacting
CC   at least one 3,4-dehydroproline residue in the polypeptide with an
CC   epoxidation reagent from a polypeptide containing at least one 3,4-
CC   epoxyproline residue. The method is used for studying the effects of non-
CC   natural amino acids on structure and function of polypeptides. The method
CC   is also useful for commercial production of collagen or mussel adhesive
CC   proteins (which are useful as bioadhesives), and for incorporating a wide
CC   variety of groups, including therapeutic ligands and biological probes,
CC   into polypeptides.
XX
SQ   Sequence 1058 AA;

Query Match   96.6%;   Score 2966;   DB 3;   Length 1058;
Best Local Similarity 95.6%;   Pred. No. 2.7e-169;

XX
PD   11-DEC-1995.
XX
PF   12-JUN-1995;   95CA-02151547.
XX
PR   10-JUN-1994;   94US-00255263.
XX
PA   (USSU) US SURGICAL CORP.
XX
PI   Gruskin EA, Espino P;
XX
DR   WPI; 1996-140144/15.
DR   N-PSDB; AAT16518.
XX
PT   Chimaeric DNA encoding protein contg. extracellular matrix protein domain
PT   - and cellular regulatory factor domain, partic. useful as osteogenic
PT   agents, also related vectors, transformed cells and chimaeric proteins.
XX
PS   Disclosure; Fig 8; 59pp; English.
XX
CC   A fusion protein (AAR89472) comprises the alpha-helical region of human
CC   collagen I(a) linked to amino acids 46-93 of human mature dermatan
CC   sulphate proteoglycan (decorin). It can be expressed in Escherichia coli
CC   transformants carrying a vector incorporating a chimeric gene (AAT16518)
CC   coding for the fusion. The decorin binds to type I collagen and thus
CC   affects Elbril formation. It inhibits the cell attachment-promoting
CC   activity of collagen and fibrinogen by binding to such molecules near
CC   their cell binding sites. The collagen moiety provides an integral
CC   substratum or scaffolding for the decorin. The fusion protein acts to
CC   reduce scarring of healing tissue
XX
SQ   Sequence 1107 AA;

Query Match   96.6%;   Score 2966; DB 2;   Length 1107;
Best Local Similarity 95.6%;   Pred. No. 2.8e-169;
Matches 520;   Conservative 21; Mismatches 3;   Indels 0;   Gaps 0;

Qy   1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE   60
          ||||||:|||||||||||:|||:||:||||:||||||||||||:||||||||||||:
Db   201 GSEGPQGVRGEPGPPGPAGAGAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ 260

Qy   61 GPGGPEGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEGKPGARGEPGPTGLP 120
          ||||||||||||:|||||:|||||||||||||||:||:||||||||||||||||||||||||
Db   261 GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEBGKRGARGEPGPTGLP   320

Qy   121 GPPGERGGPGSRGFPGADGVAGPKGPGAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
          ||||||||||||||:|||||||||||:|||:|||||||||||||||||||||||||||||||
Db   321 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT   380

Qy   181 GSPGSPGPDGKTPGPPGPAGEDGRPGPPGPPGPPGRAGEAGVMGFPGPKGAAGEPGKAGERGVP 240
          |||||:||||||||:|||||:|||||||||||:|||:||||||||||||:|||||||||
Db   381 GSPGSPGPDGKTPGPPGPAGQQGRPGPPGPPGPAGRGQAGVMGFPGPKGAAGEPGKAGERGVP 440

Qy   241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEEGPAGSPGFEGLPGPAGPPGEAGKPGEE 300
          |||||:||||||:||||:|||:|||||||||||||||||||||||||||||||:|||||
Db   441 GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQ   500

Qy   301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGADGAPGDDCAKGDAGAPGAPGSE 360
          ||||||||||||:||||||||||||||||||||:|||:||||||||||||:||||||||
Db   501 GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPPGANGAPGNDGAKGDAGAPGAPGSQ 560

Qy   361 GAPGLEGMPGERGAAGLPGPKGDRGADGPKGKADGSPGKGKDGVRGLITGPIGPPGPAGAPGDK 420
          |||||:|||||||||||||||||||||||||||||||||||||||:||||:||||||||||||
Db   561 GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGKADGSPGKGKDGVRGLITGPIGPPGPAGAPGDK 620

Qy   421 GESGPSGPAGPTGARGAPGDRGEPGEPPGPAGPAGPPGADGEPGAKGEPGDAGAKGDAGPP 480
          ||||||||||||||||||||||:|||||||:||||||||||||||||||||||||||||
Db   621 GESGPSGPAGPTGARGAPGDRGEPGEPPGPAGPAGPPGADGQPGAKGEPGDAGAKGDAGPP 680

Qy   481 GPAGPAGPPGIGDVGAPGAKGARGSAGPPGATGPGPSGDAGPPGPPGPA 540
          |||||||||::||||||||||||||:|||||:||||||||||||||||||
Db   681 GPAGPAGPPGIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPA 740

Qy   541 GKEG 544
          ||||
Db   741 GKEG 744

RESULT 5
AAY84540
ID   AAY84540 standard; protein; 1107 AA.
XX
AC   AAY84540;
XX
DT   25-JUL-2000 (first entry)
XX
DE   Amino acid sequence of a chimeric collagen 1 (alpha1)/decorin protein.
XX
KW   Extracellular matrix protein; self aggregation; hydroxylated proline;
KW   trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
KW   collagen; fibrinogen; fibronectin; post translational hydroxylation;
KW   decorin; chimera.
XX
OS   Homo sapiens.
OS   Unidentified.
OS   Chimeric.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 858
FT   /note= "Gly encoded by GCT"
XX
PN   EP992586-A2.
XX
PD   12-APR-2000.
XX
PF   07-OCT-1999;   99EP-00119184.
XX
PR   09-OCT-1998;   98US-00169768.
XX
PA   (USSU) US SURGICAL CORP.
XX
PI   Gruskin EA, Buechter DD, Zhang G, Connolly K;
XX
DR   WPI; 2000-259138/23.
DR   N-PSDB; AAA12500.
XX
PT   Production of extracellular matrix proteins containing 4-trans-
PT   hydroxyproline results in native self aggregating proteins, useful on
PT   medical implants.
XX
PS   Claim 24; Fig 18; 260pp; English.
XX
CC   The specification describes a method for producing an extracellular
CC   matrix protein or its fragment. The extracellular matrix protein is
CC   capable of self aggregating in a cell which does not ordinarily
CC   hydroxylated prolines. The method comprises optimising a nucleic acid
CC   sequence for expression in the cell by substitution of codons preferred
CC   by that cell for naturally occurring codons not preferred by the cell;
CC   incorporating the nucleic acid sequence into the cell; and contacting the
CC   cell with a hypertonic growth medium containing at least one amino acid,
CC   selected from the group consisting of trans-4-hydroxyproline and 3-
CC   hydroxyproline to allow at least one of the amino acids to be assimilated
CC   into the cell and incorporated into the extracellular matrix protein. The
CC   method may be used to make host cells assimilate and incorporate trans-4-
CC   hydroxyproline into proteins. This is especially useful in the
CC   recombinant production of proteins such as collagen, fibrinogen and
CC   fibronectin whose ability to self aggregate and produce functional
CC   proteins depends on the post translational hydroxylation of proline. The
CC   method is also useful in studying the structure and function of proline
CC   polypeptides which do not normally contain trans-4-hydroxyproline. The
CC   present sequence represents a chimeric collagen 1 (alpha1)/decorin
CC   protein, which may be produced using the method of the invention
XX
SQ   Sequence 1107 AA;

      Query Match   96.6%; Score 2966; DB 3;   Length 1107;

Left column (RESULT 6 — ADE87050)

```
Best Local Similarity   95.6%;  Pred. No. 2.8e-169;
Matches 520;  Conservative  21;  Mismatches  3;  Indels  3;  Gaps   0;

Qy     1  GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE   60
          ||||:|||||||||||||||||:||||:|||||||||:||||||||||||||||||||||:
Db   201  GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ  260

Qy    61  GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGVGVEGPPGPAGEEGKPGARGEPGPTGLP  120
          ||||||||||:|||||||||||||||||||:|:||||||||||:||:|||||||||||||
Db   261  GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGVGVQGPPGPAGEEGKRGARGEPGPTGLP  320

Qy   121  GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   321  GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT  380

Qy   181  GSPGSPGPDGKTGPPGAEDGRPGPPGPPGARGEAGVMGPPGPKGAAGEPGKAGERGVP  240
          |||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db   381  GSPGSPGPDGKTGPPGAEDGRPGPPGPPGARGQAGVMGPPGPKGAAGEPGKAGERGVP  440

Qy   241  GPPGAVGPAGKDGEAGAEGPPGPAGPAGEREGPAGSPGPAGPPGEAGKPGEE  300
          |||||||||||||||||:||||||||||||:|:|||||||:|:|:|||||:
Db   441  GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGPFQGLPGPPGPEAGKPGEQ  500

Qy   301  GVPGDLGAPGPSGARGEPGFPGERGVEGPPGRGVEGPPGPAGPPGADGAPGSE  360
          ||||||||||||||||||||||||||:|:|:||||:||||||||:|||||||||
Db   501  GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQ  560

Qy   361  GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  420
          |||||||:||||||||||||||||||||||||||:||||||||:||||||||||||||||
Db   561  GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  620

Qy   421  GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPDGAKGEPGDAGAKGDAGPP  480
          ||||||||||||||||||||||||||||||||||:||:||||||||||||||
Db   621  GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPDGAKGEPGDAGAKGDAGPP  680

Qy   481  GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAAGRVGPPGPPGPGPA  540
          |||||||||||:|||||||||||||||||||||||||||||||||||||:||||
Db   681  GPAGPAGPPGPIGNVGAPGAKGARGSAGPPGATGFPGAAAGRVGPPGPSGNAGPPGPPGPA  740

Qy   541  GKEG  544
          ||||
Db   741  GKEG  744
```

RESULT 6
ADE87050
ID   ADE87050 standard; protein; 1161 AA.
XX
AC   ADE87050;
XX
DT   29-JAN-2004 (first entry)
XX
DE   Human pancreatic cell protein sequence SeqID510.
XX
KW   neoplastic pancreatic cell; pancreatic cell; pancreatic cancer;
KW   cancer death; cytostatic; vaccine; gene therapy;
KW   non-cancerous pancreas disease; human.
XX
OS   Homo sapiens.
XX
PN   WO200306145-A2.
XX
PD   24-JUL-2003.
XX
PF   19-DEC-2002; 2002WO-US040655.
XX
PR   21-DEC-2001; 2001US-0342768P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Sun Y, Liu C;
XX
DR   WPI; 2003-587286/55.

Right column (RESULT 7 — AAR89469)

DR   N-PSDB; ADE87387.
XX
PT   New pancreatic specific nucleic acid molecule or protein for diagnosing,
PT   staging, imaging, monitoring, preventing or treating pancreatic cancer or
PT   non-cancerous disease states of the pancreas.
XX
PS   Claim 12; SEQ ID NO 510; 635pp; English.
XX
CC   This invention relates to novel nucleic acids and proteins present in
CC   normal and neoplastic pancreatic cells. Pancreatic cancer is a common
CC   cause of cancer death worldwide, therefore accurate methods of diagnosis
CC   and treatment are required. Compounds which modulate the proteins of the
CC   invention may have cytostatic activity and the protein and DNA sequences
CC   of the invention may be useful for the development of a vaccine or in
CC   gene therapy. The composition and methods are useful in diagnosing,
CC   staging, imaging, monitoring, preventing or treating pancreatic cancer
CC   and non-cancerous disease states of the pancreas. The present sequence is
CC   that of a human pancreatic protein of the invention.
XX
SQ   Sequence 1161 AA;

```
Query Match   96.6%;  Score 2966;  DB 7;  Length 1161;
Best Local Similarity   95.6%;  Pred. No. 2.9e-169;
Matches 520;  Conservative  21;  Mismatches  3;  Indels  3;  Gaps   0;

Qy     1  GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE   60
          ||||:|||||||||||||||||:||||:|||||||||:||||||||||||||||||||||:
Db    59  GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ  118

Qy    61  GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGVGVEGPPGPAGEEGKPGARGEPGPTGLP  120
          ||||||||||:|||||||||||||||||||:|:||||||||||:||:|||||||||||||
Db   119  GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGVGVQGPPGPAGEEGKRGARGEPGPTGLP  178

Qy   121  GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   179  GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT  238

Qy   181  GSPGSPGPDGKTGPPGAEDGRPGPPGPPGARGEAGVMGPPGPKGAAGEPGKAGERGVP  240
          |||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db   239  GSPGSPGPDGKTGPPGAEDGRPGPPGPPGARGQAGVMGPPGPKGAAGEPGKAGERGVP  298

Qy   241  GPPGAVGPAGKDGEAGAEGPPGPAGPAGEREGPAGSPGPAGPPGEAGKPGEE  300
          |||||||||||||||||:||||||||||||:|:|||||||:|:|:|||||:
Db   299  GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQ  358

Qy   301  GVPGDLGAPGPSGARGEPGFPGERGVEGPPGRGVEGPPGPAGPPGADGAPGSE  360
          ||||||||||||||||||||||||||:|:|:||||:||||||||:|||||||||
Db   359  GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQ  418

Qy   361  GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  420
          |||||||:||||||||||||||||||||||||||:||||||||:||||||||||||||||
Db   419  GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  478

Qy   421  GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPDGAKGEPGDAGAKGDAGPP  480
          ||||||||||||||||||||||||||||||||||:||:||||||||||||||
Db   479  GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPDGAKGEPGDAGAKGDAGPP  538

Qy   481  GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAAGRVGPPGPPGPGPA  540
          |||||||||||:|||||||||||||||||||||||||||||||||||||:||||
Db   539  GPAGPAGPPGPIGNVGAPGAKGARGSAGPPGATGFPGAAAGRVGPPGPSGNAGPPGPPGPA  598

Qy   541  GKEG  544
          ||||
Db   599  GKEG  602
```

RESULT 7
AAR89469
ID   AAR89469 standard; protein; 1169 AA.
XX
AC   AAR89469;
XX
DT   01-OCT-1996 (first entry)

```
XX
DE   Collagen/BMP-2B fusion protein.
XX
KW   Bone morphogenic protein 2B; BMP-2B; collagen IA; osteogenesis;
KW   fusion protein.
XX
OS   Synthetic.
XX
FH   Key              Location/Qualifiers
FT   Domain           1..1057
FT                    /label= Collagen-IA
FT                    /note= "collagen IA alpha-helical domain"
FT   Misc-difference  887
FT                    /note= "unidentified amino acid"
FT   Misc-difference  890
FT                    /note= "unidentified amino acid"
FT   Peptide          1058..1059
FT                    /label= Linker_peptide
FT   Domain           1060..1169
FT                    /label= BMP-2B
FT                    /note= "human mature BMP-2B"
XX
PN   CA2151547-A.
XX
PD   11-DEC-1995.
XX
PF   12-JUN-1995;  95CA-02151547.
XX
PR   10-JUN-1994;  94US-00259263.
XX
PA   (USSU ) US SURGICAL CORP.
XX
PI   Gruskin EA, Espino P;
XX
DR   WPI; 1996-140144/15.
DR   N-PSDB; AAT16515.
XX
CC   Chimaeric DNA encoding protein contg. extracellular matrix protein domain
CC   - and cellular regulatory factor domain, partic. useful as osteogenic
CC   agents, also related vectors, transformed cells and chimaeric proteins.
XX
PS   Disclosure; Fig 5; 59pp; English.
XX
CC   A fusion protein (AAR89469) comprises the alpha-helical region of human
CC   collagen I(a) linked to the human mature bone morphogenic protein 2B
CC   (BMP2B). It can be expressed in Escherichia coli transformants carrying a
CC   vector incorporating a chimeric gene (AAT16515) coding for the fusion.
CC   The BMP moiety induces osteogenesis, while the collagen moiety provides
CC   an integral substratum or scaffolding for the BMP and cells involved in
CC   reconstruction and growth. The fusion protein provides sustained release
CC   and delivery of BMP to a target tissue
XX
SQ   Sequence 1169 AA;

Query Match             96.6%;  Score 2966;  DB 2;  Length 1169;
Best Local Similarity   95.6%;  Pred. No. 2.9e-169;
Matches 520;  Conservative 21;  Mismatches  3;  Indels  0;  Gaps  0;

Qy     1  GSGGPEGVRGEPGPGPPGAAGPGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE   60
          ||||||||| ::| ||||||||||||||| ||||||:||||||||||:||||||||||:
Db   201  GSGEGPQGVRGEPGPGPPGAAGPGDPGAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ  260

Qy    61  GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGVGVEGPPGPAGEEGKPGARGEPGPTGLP  120
          ||||||||:||| ||||:|||||||||| |||||:|||||||||||||||||||||||||
Db   261  GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGVGVQGPPGPAGEEGKRGARGEPGPTGLP  320

Qy   121  GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGERGSPGEAGRPGEAGLPGAKGLT  180
          |||||||:|||||||||||||||||:||||||||||||||:|||:||||||||||||||||
Db   321  GPPGERGDPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT  380

Qy   181  GSPCSGPDGKTGPPGPPAGQDGRPGPGPPGARGEAGVMGFPCPKGAAGEPGKAGERGVP  240
          |||||||||||||||||||| |||:|||||||||||||||||||||||||||||||||||
Db   381  GSPGSGPDGKTGPPGPPAGQDGRPGPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGVP  440
```

```
Qy   241  GPPGAVGPAGKDGEAGABGPPGPAGPAGERGERGPAGSPGFEGLPGPAGPPGEAGKPGEE   300
          ||||||||||||||||||:|||||:||:||||||||||||:|:|:||||||:||||:|||:
Db   441  GPPGAVGPAGKDGEAGAQGPPGPAGERGEQGPAGSPGFQLPGPAGPPGEAGKPGEQ   500

Qy   301  GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPGADGAPGDDGAKGDAGAPGSE   360
          |||||||||||||||||:||||||||:||||||||||||||||||:||||||||||:
Db   501  GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGSQ   560

Qy   361  GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGPAGDK  420
          ||| |||||||||||||:||||||||||||||||:|||||||||||||:||||||||:|||
Db   561  GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGPAPGDK  620

Qy   421  GESGPSGAGPTGARGAPGDRGEPGPPGPAGFAGPAGPGDGEPGAKGDAGGPP   480
          ||||||||||||||||||:||||:|||||||| |:|||||:||||||||||||||
Db   621  GESGPSGPAGPTGARGAPDRGEPGFPGPAGFAGPGDGQPGAKGEPGDAGAKGDAGPP  680

Qy   481  GPAGPAGPPGPIGVGAPGAKGARGSAGPGDATGFPGAAGRVGPPGPSGDAGPGPPGPGPA  540
          ||||||| |||||:||||||||||||||||:|:||||||||||||||:|:||||:|||||
Db   681  GPAGPAGPPGPIGNVGAPGAKGARGSAGPGDATGFPGAAGRVGPPGPSGNAGPPGPPGPA  740

Qy   541  GKEG  544
          ||||
Db   741  GKEG  744
```

```
RESULT 8
AAY84537
ID   AAY84537 standard; protein; 1169 AA.
XX
AC   AAY84537;
XX
DT   25-JUL-2000  (first entry)
XX
DE   Amino acid sequence of a chimeric collagen 1 (alpha1)/BMP-2B protein.
XX
KW   Extracellular matrix protein; self aggregation; hydroxylated proline;
KW   trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
KW   collagen; fibrinogen; fibronectin; post translational hydroxylation;
KW   ss. bone morphogenic protein; BMP-2B; chimera.
XX
OS   Homo sapiens.
OS   Unidentified.
OS   Chimeric.
XX
FH   Key              Location/Qualifiers
FT   Misc-difference  677
FT                    /note= "Ala encoded by G"
FT   Misc-difference  887
FT                    /note= "unspecified amino acid encoded by CT"
FT   Misc-difference  890
FT                    /note= "unspecified amino acid encoded by CT"
XX
PN   EP992586-A2.
XX
PD   12-APR-2000.
XX
PF   07-OCT-1999;  99EP-00119184.
XX
PR   09-OCT-1998;  98US-00169768.
XX
PA   (USSU ) US SURGICAL CORP.
XX
PI   Gruskin EA,  Buechter DD,  Zhang G,  Connolly K;
XX
DR   WPI; 2000-259138/23.
DR   N-PSDB; AAA12497.
XX
PT   Production of extracellular matrix proteins containing 4-trans-
PT   hydroxyproline results in native self aggregating proteins, useful on
PT   medical implants.
XX
PS   Claim 22; Fig 13; 260pp; English.
```

XX
CC The specification describes a method for producing an extracellular
CC matrix protein or its fragment. The extracellular matrix protein is
CC capable of self aggregating in a cell which does not ordinarily
CC hydroxylated prolines. The method comprises optimising a nucleic acid
CC sequence for expression in the cell by substitution of codons preferred
CC by that cell for naturally occurring codons not preferred by the cell;
CC incorporating the nucleic acid sequence into the cell; and contacting the
CC cell with a hypertonic growth medium containing at least one amino acid,
CC selected from the group consisting of trans-4-hydroxyproline and 3-
CC hydroxyproline to allow at least one of the amino acids to be assimilated
CC into the cell and incorporated into the extracellular matrix protein. The
CC method may be used to make host cells assimilate and incorporate trans-4-
CC hydroxyproline into proteins. This is especially useful in the
CC recombinant production of proteins such as collagen, fibrinogen and
CC fibronectin whose ability to self aggregate and produce functional
CC proteins depends on the post translational hydroxylation of proline. The
CC method is also useful in studying the structure and function of
CC polypeptides which do not normally contain trans-4-hydroxyproline. The
CC present sequence represents a chimeric collagen 1 (alpha1)/bone
CC morphogenic protein-2B (bmp-2b) protein, which may be produced using the
CC method of the invention
XX
SQ Sequence 1169 AA;

Query Match   96.6%;   Score 2966; DB 3;   Length 1169;
Best Local Similarity 95.6%;   Pred. No. 2.9e-169;
Matches 520; Conservative   21; Mismatches   3;   Indels   0;   Gaps   0;

Qy   1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE   60
           ||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db   201 GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ 260

Qy   61 GPGGPPGPKDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEGKPGARGEPGPTGLP   120
           ||||||||||:|||||||||||||||||||||:||||||||||||||||||||||||||
Db   261 GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEGEKGARGEPGPTGLP 320

Qy   121 GPPGERGGPGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   321 GPPGERGGPGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 380

Qy   181 GSPGSPGPDGKTGPPGPAGPAGEDGRDGPPGPPGPPGARGEAGVMGFPGPKGEAAGEKGAGERGVP 240
           |||||||||||||||||||||::|||||||||||||||||||||::|||||||||||||||||||
Db   381 GSPGSPGPDGKTGPPGPAGQDPGRPGPPGPPGPPGARGQAGVMGFPGPKGEAAGEPGKAGERGVP 440

Qy   241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEEGPAGSPGFEGLPGPAGPPGEAGKPGEE 300
           ||||||||||||||||||||||||||||||||||:|||::||||||||||||||||||::
Db   441 GPPGAVGPAGKDGEAGAQGPPGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQ 500

Qy   301 GVPGDLGAPGSGARGEPGFPGERGVEGPPGPAGPPGADGAPGDDGAKGDAGAPGSE 360
           ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db   501 GVPGDLGAPGSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQ 560

Qy   361 GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGKDGVRGLTGPIGPPGPAGPAPGDK 420
           ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db   561 GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGKADGSPGKDGVRGLTGPIGPPGPAGPAPGDK 620

Qy   421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPPGADGEPGAKGEPGDAGAKGDAGPP 480
           ||||||||||||||||||||||||:::|||||||||||||||||||||||||:||||||||
Db   621 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPPGPAGPGPGAPGPGADGQPGAKGEPGDAGAKGDAGPP 680

Qy   481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPGPGATGPGAAGRVGPPGPSGDAGPPGPPGPPA 540
           ||||||||||:||||||:|||||||||||||||||||||||:||||||:|||:||:|||||
Db   681 GPAGPAGPPGPIGNVGAPGAKGARGSAGPGPGATGPGAAGRVGPPGPSGNAGPPGPPGPA 740

Qy   541 GKEG 544
           ||||
Db   741 GKEG 744

RESULT 9
AAR89470
ID AAR89470 standard; protein; 1171 AA.

XX
AC AAR89470;
XX
DT 01-OCT-1996 (first entry)
XX
DE Collagen/TGF-beta-1 fusion protein.
XX
KW Transforming growth factor; TGF-beta-1; collagen IA; osteogenesis;
KW bone formation; tissue repair; fusion protein.
XX
OS Synthetic.
XX
FH Key   Location/Qualifiers
FT Domain   1..1057
FT   /label= Collagen-IA
FT   /note= "collagen IA alpha-helical domain"
FT Misc-difference 887
FT   /note= "unidentified amino acid"
FT Misc-difference 890
FT   /note= "unidentified amino acid"
FT Peptide   1058..1059
FT   /label= Linker_peptide
FT Domain   1060..1171
FT   /label= TGF-beta-1
FT   /note= "human mature TGF-beta-1"
XX
PN CA2151547-A.
XX
PD 11-DEC-1995.
XX
PF 12-JUN-1995;   95CA-02151547.
XX
PR 10-JUN-1994;   94US-00259263.
XX
PA (USSU) US SURGICAL CORP.
XX
PI Gruskin EA, Espino P;
XX
DR WPI; 1996-140144/15.
DR N-PSDB; AAT16516.
XX
PT Chimaeric DNA encoding protein contg. extracellular matrix protein domain
PT - and cellular regulatory factor domain, partic. useful as osteogenic
PT agents, also related vectors, transformed cells and chimaeric proteins.
XX
PS Disclosure; Fig 6; 59pp; English.
XX
CC A fusion protein (AAR89470) comprises the alpha-helical region of human
CC collagen I(a) linked to the human mature transforming growth factor beta-
CC 1 (TGF-beta-1). It can be expressed in Escherichia coli transformants
CC carrying a vector incorporating a chimeric gene (AAT16516) coding for the
CC fusion. The TGF-beta- moiety increases efficacy of the body's normal soft
CC tissue repair response and also induces osteogenesis. The collagen moiety
CC provides an integral substratum or scaffolding for the TGF and cells
CC involved in reconstruction and growth. The fusion protein provides
CC sustained release and delivery of TGF-beta-1 to a target tissue
XX
SQ Sequence 1171 AA;

Query Match   96.6%;   Score 2966; DB 2;   Length 1171;
Best Local Similarity 95.6%;   Pred. No. 2.9e-169;
Matches 520; Conservative   21; Mismatches   3;   Indels   0;   Gaps   0;

Qy   1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE   60
           ||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db   201 GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ 260

Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEGKPGARGEPGPTGLP 120
           ||||||||||:|||||||||||||||||||||:||||||||||||||||||||||||||
Db   261 GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEGEKGARGEPGPTGLP 320

Qy   121 GPPGERGGPGPGSRGFPGADGVAGPKGSPGAGPKGSPGEAGRPGEAGLPGAKGLT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   261 GPPGERGGPGPGSRGFPGADGVAGPKGSPGAGPKGSPGEAGRPGEAGLPGAKGLT 320

Db   321 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSSPGEAGRPGEAGLPGAKGLT 380
```
              ||||||||||||:|||||:|:||||||||||||||||||||||
```
Qy   181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGPPGARGEAGVMGFPGPKGAAGEPGKAGERGVP 240
```
              ||||||||||||||||||||:|||||||||||||||||||||
```
Db   381 GSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGVP 440

Qy   241 GPPGAVGPAGKDGEAGAEGPPGPAGAGEREGEPAGSPGFEGLPGPAGPPGEAGKPGEE 300
```
              |||||||||||||:|||:||||:||||||||||||||||||||
```
Db   441 GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGQLPGPAGPPGEAGKPGEQ 500

Qy   301 GVPGDLGAPGSGARGEPGFPGERGVPGDLGAPGPPGPAGPPGDADGAPGDDGAKGDAGAPGAPGSE 360
```
              ||||||||||||:|:|||:||||||||||||||||||:|:|:|
```
Db   501 GVPGDLGAPGPSGARGERGFPGBERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQ 560

Qy   361 GAPGLEMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
```
              |||||||||||||:|||||||||||||||||||||||||||||
```
Db   561 GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 620

Qy   421 GESGPSGPAGPTGARGARGPGADGREGEPGPPGPAGFAGPGPGPAGEPGDAGAKGDAGPP 480
```
              |||||||||||||||||:|||||||||||||||||||||||||
```
Db   621 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPGPGADGQPGAKGEPGDAGAKGDAGPP 680

Qy   481 GPAGPAGPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPPGPPGPA 540
```
              ||||||||||:|:|||||||||||||||||||||||||||||
```
Db   681 GPAGPAGPGPIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPA 740

Qy   541 GKEG 544
```
              ||||
```
Db   741 GKEG 744

RESULT 10
AAY84538

ID   AAY84538 standard; protein; 1171 AA.
XX
AC   AAY84538;
XX
DT   25-JUL-2000 (first entry)
XX
DE   A chimeric collagen 1 (alpha1)/TGF-beta1 protein.
XX
KW   Extracellular matrix protein; self aggregation; hydroxylated proline;
KW   trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
KW   collagen; fibrinogen; fibronectin; post translational hydroxylation;
KW   ss. transforming growth factor-beta1; TGF-beta1; chimera.
XX
OS   Homo sapiens.
OS   Unidentified.
OS   Chimeric.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 858
FT   /note= "Gly encoded by GCT"
XX
PN   EP992586-A2.
XX
PD   12-APR-2000.
XX
PF   07-OCT-1999; 99EP-00119184.
XX
PR   09-OCT-1998; 98US-00169768.
XX
PA   (USSU) US SURGICAL CORP.
XX
PI   Gruskin EA, Buechter DD, Zhang G, Connolly K;
XX
DR   WPI; 2000-259138/23.
DR   N-PSDB; AAA12498.
XX
PT   Production of extracellular matrix proteins containing 4-trans-
PT   hydroxyproline results in native self aggregating proteins, useful on
PT   medical implants.
XX

PS   Claim 23; Fig 15; 260pp; English.
XX
CC   The specification describes a method for producing an extracellular
CC   matrix protein or its fragment. The extracellular matrix protein is
CC   capable of self aggregating in a cell which does not ordinarily
CC   hydroxylated prolines. The method comprises optimising a nucleic acid
CC   sequence for expression in the cell by substitution of codons preferred
CC   by that cell for naturally occurring codons not preferred by the cell;
CC   incorporating the nucleic acid sequence into the cell; and contacting the
CC   cell with a hypertonic growth medium containing at least one amino acid,
CC   selected from the group consisting of trans-4-hydroxyproline and 3-
CC   hydroxyproline to allow at least one of the amino acids to be assimilated
CC   into the cell and incorporated into the extracellular matrix protein. The
CC   method may be used to make host cells assimilate and incorporate trans-4-
CC   hydroxyproline into proteins. This is especially useful in the
CC   recombinant production of proteins such as collagen, fibrinogen and
CC   fibronectin whose ability to self aggregate and produce functional
CC   proteins depends on the post translational hydroxylation of proline. The
CC   method is also useful in studying the structure and function of.
CC   polypeptides which do not normally contain trans-4-hydroxyproline. The
CC   present sequence represents chimeric collagen 1 (alpha1)/transforming
CC   growth factor-beta1 (TGF-beta1) protein, which may be produced using the
CC   method of the invention
XX
SQ   Sequence 1171 AA;

Query Match   96.6%;   Score 2966;   DB 3;   Length 1171;
Best Local Similarity   95.6%;   Pred. No. 2.9e-169;
Matches 520;   Conservative 21;   Mismatches 3;   Indels 0;   Gaps 0;

Qy   1 GSEGPEGVRGEPGPPGPAGAGAPGPGAGAGEPGAKGADGAPGIAGAPGFPGARGPSGPE 60
```
              ||||::||:|||||||||||:||||||||::||:||||||||||||||||||
```
Db   201 GSEGPQGVRGEPGPPGPAGAGAPAGNPGADGQPGADGAPGTAGAPGFPGARGPSGPQ 260

Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEEGKPGARGEPGPTGLP 120
```
              ||||||||||||:|||:||||||||||||||||||||||||||||||||||
```
Db   261 GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEEGKRGARGEPGPTGLP 320

Qy   121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSSPGEAGRPGEAGLPGAKGLT 180
```
              |||||||||||||:|||||:|:||||||||||||||||||||:||||||||
```
Db   321 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSSPGEAGRPGEAGLPGAKGLT 380

Qy   181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGPPGARGEAGVMGFPGPKGAAGEPGKAGERGVP 240
```
              ||||||||||||||||||||:||||||||||||||||||||||
```
Db   381 GSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGVP 440

Qy   241 GPPGAVGPAGKDGEAGAEGPPGPAGAGEREGEPAGSPGFEGLPGPAGPPGEAGKPGEE 300
```
              |||||||||||||:|||:||||:||||||||||||||||||||
```
Db   441 GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQ 500

Qy   301 GVPGDLGAPGSGARGEPGEPGFPGERGVEGPPGPAGPPGDADGAPGDDGAKGDAGAPGAPGSE 360
```
              ||||||||||:||||||:||||||||||||||:|:|:||:||
```
Db   501 GVPGDLGAPGPSGARGERGFPGBERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQ 560

Qy   361 GAPGLEMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
```
              |||||||||||||:|||||||||||||||||||||||||||||
```
Db   561 GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 620

Qy   421 GESGPSGPAGPTGARGARGSAGPPGATGFPGAAGRVGPPGPSGDAGP 480
```
              ||||||||||||||||||:||||||||||||||||||||||||
```
Db   621 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPGPGADGQPGAKGEPGDAGAKGDAGPP 680

Qy   481 GPAGPAGPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPPGPPGPA 540
```
              ||||||||||:|:|||||||||||||||||||||||||||||
```
Db   681 GPAGPAGPGPIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPA 740

Qy   541 GKEG 544
```
              ||||
```
Db   741 GKEG 744

RESULT 11
ADE87057

```
ID   ADE87057 standard; protein; 1211 AA.
XX
AC   ADE87057;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human pancreatic cell protein sequence SeqID517.
XX
KW   neoplastic pancreatic cell; pancreatic cell; pancreatic cancer;
KW   cancer death; cytostatic; vaccine; gene therapy;
KW   non-cancerous pancreas disease; human.
XX
OS   Homo sapiens.
XX
PN   WO2003060145-A2.
XX
PD   24-JUL-2003.
XX
PR   19-DEC-2002; 2002WO-US040655.
XX
PR   21-DEC-2001; 2001US-0342768P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Sun Y, Liu C;
XX
DR   WPI; 2003-587286/55.
DR   N-PSDB; ADE87397.
XX
PT   New pancreatic specific nucleic acid molecule or protein for diagnosing,
PT   staging, imaging, monitoring, preventing or treating pancreatic cancer or
PT   non-cancerous disease states of the pancreas.
XX
PS   Claim 12; SEQ ID NO 517; 635pp; English.
XX
CC   This invention relates to novel nucleic acids and proteins present in
CC   normal and neoplastic pancreatic cells. Pancreatic cancer is a common
CC   cause of cancer death worldwide, therefore accurate methods of diagnosis
CC   and treatment are required. Compounds which modulate the proteins of the
CC   invention may have cytostatic activity and the protein and DNA sequences
CC   of the invention may be useful for the development of a vaccine or in
CC   gene therapy. The composition and methods are useful in diagnosing,
CC   staging, imaging, monitoring, preventing or treating pancreatic cancer
CC   and non-cancerous disease states of the pancreas. The present sequence is
CC   that of a human pancreatic protein of the invention.
XX
SQ   Sequence 1211 AA;

Query Match             96.6%;  Score 2966; DB 7; Length 1211;
Best Local Similarity   95.6%;  Pred. No. 3e-169;
Matches  520;  Conservative   21; Mismatches    3; Indels    0; Gaps    0;

Qy     1 GSEGPGVRGEPGPPGPAGPAGPPGPAGDGEPGAKGADGAPGIAGAPGPGPGARGPSGPE 60
           ||||| ||||||||||||| ||||||  |||||||| ||||||||||||||||||||
Db   362 GSEGPQGVRGEPGPPGPAGPAGPAGNPGADGQPGAKGANGAPGIAGAPGPGPGARGPSGPQ 421

Qy    61 GPGGPPGPKGDSGEPGAPGSKDTGAKGEPGPGVGVEGPGPPGPAGEGKPGARGEPGPTGLP 120
           ||||||||| :||||||||||||||||||||||| |||||||||||||:||||||||||
Db   422 GPGGPPGPKGNSGEPGAPGSKDTGAKGEPGPGVGVQGPGPPGPAGEGKRGARGEPGPTGLP 481

Qy   121 GPPGERGPGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   482 GPPGERGPGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 541

Qy   181 GSPGSPGPDGKTGPGPGPAGEDGRPGPPGPPGPPGARGEAGVMGFPGPGPGKGAAGBPGKAGBRGVP 240
           |||||||||||||| |||| ||||||||||||||||| |||||||||||||||||||||||||||
Db   542 GSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGPPGARGQAGVMGFPGPGPGKGAAGBPGKAGBRGVP 601

Qy   241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEEGPAGSPGFEGLPGPAGPPGEAGKPGEE 300
           |||||||||||||||| |||  ||||||||| ||||||||| |||||||||||||||||:
Db   602 GPPGAVGPAGKDGEAGAQGPGPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQ 661

Qy   301 GVPGDLGAPGPSGARGEPGFPGERGVTGPGPPGPADGAPGDDGAKGDAGAPGAPGAPGSE 360
```

```
RESULT 12
ADE87062
ID   ADE87062 standard; protein; 1226 AA.
XX
AC   ADE87062;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human pancreatic cell protein sequence SeqID522.
XX
KW   neoplastic pancreatic cell; pancreatic cell; pancreatic cancer;
KW   cancer death; cytostatic; vaccine; gene therapy;
KW   non-cancerous pancreas disease; human.
XX
OS   Homo sapiens.
XX
PN   WO2003060145-A2.
XX
PD   24-JUL-2003.
XX
PR   19-DEC-2002; 2002WO-US040655.
XX
PR   21-DEC-2001; 2001US-0342768P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Sun Y, Liu C;
XX
DR   WPI; 2003-587286/55.
DR   N-PSDB; ADE87403.
XX
PT   New pancreatic specific nucleic acid molecule or protein for diagnosing,
PT   staging, imaging, monitoring, preventing or treating pancreatic cancer or
PT   non-cancerous disease states of the pancreas.
XX
PS   Claim 12; SEQ ID NO 522; 635pp; English.
XX
CC   This invention relates to novel nucleic acids and proteins present in
CC   normal and neoplastic pancreatic cells. Pancreatic cancer is a common
CC   cause of cancer death worldwide, therefore accurate methods of diagnosis
CC   and treatment are required. Compounds which modulate the proteins of the
CC   invention may have cytostatic activity and the protein and DNA sequences
CC   of the invention may be useful for the development of a vaccine or in
CC   gene therapy. The composition and methods are useful in diagnosing,
CC   staging, imaging, monitoring, preventing or treating pancreatic cancer
CC   and non-cancerous disease states of the pancreas. The present sequence is
CC   that of a human pancreatic protein of the invention.
XX
SQ   Sequence 1226 AA;

Query Match             96.6%;  Score 2966; DB 7; Length 1226;
Best Local Similarity   95.6%;  Pred. No. 3e-169;
Matches  520;  Conservative   21; Mismatches    3; Indels    0; Gaps    0;

Db   662 GVPGDLGAPGPSGARGERGFPGERGVQGPGPPAGPGERGVGQPGPPGAGPNGDGAKGDAGAPGAGSQ 721
           ||||||||||||||||| ||||||| |: ||||||::|| :|||||| |||||||||||||||
Qy   361 GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPTGPPGPPAGAPGDK 420
           |||| ||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   722 GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPPAGAPGDK 781

Qy   421 GESGPSGPAGPTGARGAPGDRGEPGPPGAGPAGPPGADGEPGAKGEPGDAGAKGDAGPP 480
           ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   782 GESGPSGPAGPTGARGAPGDRGEPGPPGAGPAGPPGADGQPGAKGEPGDAGAKGDAGPP 841

Qy   481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPPGPPGPA 540
           ||||||||||||| |||||||||||||||||||||||||||||||||:||||||||||
Db   842 GPAGPAGPPGPIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPA 901

Qy   541 GKEG 544
           ||||
Db   902 GKEG 905
```

```
XX
PF   12-JUN-1995;    95CA-02151547.
XX
PR   10-JUN-1994;    94US-00259263.
XX
PA   (USSU ) US SURGICAL CORP.
XX
PI   Gruskin EA, Espino P;
XX
DR   WPI; 1996-140144/15.
DR   N-PSDB; AAT16517.
XX
PT   Chimaeric DNA encoding protein contg. extracellular matrix protein domain
PT   - and cellular regulatory factor domain, partic. useful as osteogenic
PT   agents, also related vectors, transformed cells and chimeric proteins.
XX
PS   Disclosure; Fig 7; 59pp; English.
XX
CC   A fusion protein (AAR89471) comprises the alpha-helical region of human
CC   collagen I(a) linked to human dermatan sulphate proteoglycan (decorin).
CC   It can be expressed in Escherichia coli transformants carrying a vector
CC   incorporating a chimeric gene (AAT16517) coding for the fusion. The
CC   decorin binds to type I collagen and thus affects fibril formation. It
CC   inhibits the cell attachment-promoting activity of collagen and
CC   fibringen by binding to such molecules near their cell binding sites.
CC   The collagen moiety provides an integral substratum or scaffolding for
CC   the decorin. The fusion protein acts to reduce scarring of healing tissue
XX
SQ   Sequence 1388 AA;
```

Query Match   96.6%; Score 2966; DB 2; Length 1388;
Best Local Similarity   95.6%; Pred. No. 3.4e-169;
Matches 520; Conservative   21; Mismatches 3; Indels   0; Gaps   0;

```
Qy      1  GSEGPEGVRGEPGPPGPAGAAGPAGDGEPGADGEPGAKGDAGPGIAGAPGFPGPARGPSGPE   60
           |||||:|||||||||||||||||||||||  :|||: :|||||:|:| |||||||||| ||||:|:
Db    201  GSEGPQGVRGEPGPPGPAGAAGPAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ  260

Qy     61  GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPVGVQGPPGPAGEGEKPGARGEPGPTGLP  120
           ||||||||||:||||||||||||||||||||||||:||||||||||||||||||||||:|||||||||||
Db    261  GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPVGVQGPPGPAGEGEKRGARGEPGPTGLP  320

Qy    121  GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT  180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321  GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT  380

Qy    181  GVPGDLGAPGSGARGEPGPPGERGVEGPPGPAGEDGRPGPPGPGPPGPGARGEAGVMGFPGPKGAAGEPGKAGERGVP  240
           ||||||||||||||||||||||||:||||||||::||||||||||   |||:||||||||||||||||||||||||||
Db    381  GSPGSPGPDGKTGPPGPAGQDGRPGPPGPGARGQAGVMGFPGPKGAAGEPGKAGERGVP  440

Qy    241  GPPGAVGPAGKDGEAGAEGPPGPAGERGEGPAGSPGFEGLPGPAGPPGPAGEGKPGEE  300
           |||||||||||||||||:||||||||||:||||||||:||||||||||||||||||:|
Db    441  GPPGAVGPAGKDGEAGAQGPPGPAGERGEQGPAGSPGFQGLPGPAGPPGPAGEGKPGEQ  500

Qy    301  GVPGDLGAPGSGARGEPGPPGERGVEGPPGPAGRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  360
           ||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db    501  GVPGDLGAPGSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQ  560

Qy    361  GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  420
           ||||||:||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db    561  GAPGLQGMPGRGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  620

Qy    421  GESGPSGPAGPTGARGAPGDGKAGARGSAGPGDVGAPGAKGARGSAGPPGPATGFPGAAGKPGEE  480
           |||||||||||||||||||||:|||||||||:||||||||||||||||||||||||||||:|:||
Db    621  GESGPSGPAGPTGARGAPGDGRGEPGPPGPAGPAGFAGPPGADGQPGADGAKGDAGPP  680

Qy    481  GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGPATGFPGAAGRVGPPGPSDGAPGGPPGPA  540
           |||||||||||||:||||||||||||||||||||||||||||||||||||:|:|:||||||
Db    681  GPAGPAGPPGPIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPGPA  740

Qy    541  GKEG  544
           ||||
Db    741  GKEG  744
```

Second alignment (RESULT continuation):

```
Qy     *  GSEGPEGVRGEPGPPGPAGAAGPAGDGEPGADGEPGAKGDAGPGIAGAPGFPGPARGPSGPE   60
Db    362  GSEGPQGVRGEPGPPGPAGAAGPAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ  421

Qy     61  GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPVGVQGPPGPAGEGEKPGARGEPGPTGLP  120
Db    422  GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPVGVQGPPGPAGEGEKRGARGEPGPTGLP  481

Qy    121  GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT  180
Db    482  GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT  541

Qy    181  GVPGDLGAPGSGARGEPGPPGERGVEGPPGPAGEDGRPGPPGPGPPGPGARGEAGVMGFPGPKGAAGEPGKAGERGVP  240
Db    542  GSPGSPGPDGKTGPPGPAGQDGRPGPPGPGARGQAGVMGFPGPKGAAGEPGKAGERGVP  601

Qy    241  GPPGAVGPAGKDGEAGAEGPPGPAGERGEGPAGSPGFEGLPGPAGPPGPAGEGKPGEE  300
Db    602  GPPGAVGPAGKDGEAGAQGPPGPAGERGEQGPAGSPGFQGLPGPAGPPGPAGEGKPGEQ  661

Qy    301  GVPGDLGAPGSGARGEPGPPGERGVEGPPGPAGRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  360
Db    662  GVPGDLGAPGSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQ  721

Qy    361  GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  420
Db    722  GAPGLQGMPGRGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  781

Qy    421  GESGPSGPAGPTGARGAPGDGKAGARGSAGPGDVGAPGAKGARGSAGPPGPATGFPGAAGKPGEE  480
Db    782  GESGPSGPAGPSGAGPPGARGSAGPGATGFPGATGFPGAAGRVGPPGPSDGAPGGPPGPPGPA  841

Qy    481  GPAGPAGPPGPIGDVGAPGAKGARGSAGPGATGFPGAAGRVGPPGPSDGAPGGPPGPPGPA  540
Db    842  GPAGPAGPPGPIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPGPA  901

Qy    541  GKEG  544
Db    902  GKEG  905
```

RESULT 13
AAR89471

```
ID   AAR89471 standard; protein; 1388 AA.
XX
AC   AAR89471;
XX
DT   01-OCT-1996 (first entry)
XX
DE   Collagen/decorin fusion protein.
XX
KW   Transforming growth factor; TGP-beta-1; collagen IA; osteogenesis;
KW   bone formation; tissue repair; fusion protein.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..1057
FT                   /label= Collagen-IA
FT                   /note= "collagen IA alpha-helical domain"
FT   Misc-difference 887
FT                   /note= "unidentified amino acid"
FT   Misc-difference 890
FT                   /note= "unidentified amino acid"
FT   Peptide         1058..1059
FT                   /label= Linker_peptide
FT   Domain          1060..1388
FT                   /label= Decorin
XX
PN   CA2151547-A.
XX
PD   11-DEC-1995.
```

```
RESULT 14
AAY56800
ID   AAY56800 standard; protein; 1411 AA.
XX
AC   AAY56800;
XX
DT   27-MAR-2000  (first entry)
XX
DE   Human preproalpha 1 (I) collagen.
XX
KW   Fibrillar collagen; C propeptide; SSAD; telopeptide; gelatin;
KW   sequence selection and alignment domain; prosthetic implant; foodstuff;
KW   medicine; type I collagen; human.
XX
OS   Homo sapiens.
XX
FN   EP967226-A2.
XX
PD   29-DEC-1999.
XX
PF   04-MAY-1999;  99EP-00303470.
XX
PR   08-MAY-1998;  98US-0084828P.
PR   10-APR-1999;  99US-00289578.
XX
PA   (COHE-) COHESION TECHNOLOGIES INC.
XX
PI   Olsen DR,  Hitzeman RA,  Chisholm GE;
XX
DR   WPI; 2000-074666/07.
XX
PT   New method for production of fibrillar collagen, useful for preparing
PT   telopeptide collagen fibrils and gelatin.
XX
PS   Example 1; Fig 3A-B; 30pp; English.
XX
CC   The invention provides a method for the production of fibrillar collagen.
CC   The method comprises: (a) culturing a recombinant host cell comprising a
CC   DNA encoding a fibrillar collagen monomer lacking a C propeptide SSAD
CC   (sequence selection and alignment domain); and (b) producing the
CC   fibrillar collagen. The methods are used to produce fibrillar collagen,
CC   from which telopeptide collagen fibrils can be derived. Host cells,
CC   comprising DNA encoding a collagen monomer lacking SSAD or N propeptide
CC   is used to produce gelatin. Collagen is used in biological research as a
CC   substrate for in vitro cell culture and as a component of biocompatible
CC   materials for use in prosthetic implants, sustained drug release
CC   matrices, artificial skin and wound dressing and healing devices. Gelatin
CC   is particularly useful for foodstuffs and medicine, for coating tablets
CC   and making capsules. The methods, comprising the use of collagen monomers
CC   lacking the N and/or C propeptides, result in a large increase in the
CC   production of type I collagen. The present sequence represents the human
CC   preproalpha 1 (I) collagen (GenBank Accn no: AF017178)
XX
SQ   Sequence 1411 AA;

Query Match             96.6%;  Score 2966;  DB 3;  Length 1411;
Best Local Similarity   95.6%;  Pred. No. 3.4e-169;
Matches 520;  Conservative  21;  Mismatches  3;  Indels  0;  Gaps  0;

Qy      1 GSEGPEGVRGEPGPPGPGPGPAGAAGPPAGDPAGDGEPGCAKGADGAPGIAGAPGFPGARGPSGPE  60
          ||||||| |:|||| |||||||||||||||||||||:||| ||||||:|||||||||||| |||||
Db    359 GSSEGPQGVRGEPGPPGPAGAAGPPAGNPAGDGEPGCAKGANGAPGIAGAPGFPGARGPSGPQ 418

Qy     61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGVVEGPGPGVGEPGVGQGPGPGPAGEEGKPGARGEPGPTGLP 120
          ||||| |||||:|||||||||||||||:|||| |||||||| |||||||||||||||||||:|||||||||||
Db    419 GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGVGVQGPGVGQGPGPGPAGEEGKRGARGEPGPTGLP 478

Qy    121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
          ||||||| ||||| |||||||||||||||||||:||||||||||||||| || |||||||
Db    479 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 538
```

```
                        181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGPPGARGEAGVMGFPGPGPKGAAGEPGKAGERGVP 240
                            |||||||||||||||||||||:||:|||||||||:||||||:||||||||||||||||||||||
                        539 GSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPGPKGAAGEPGKAGERGVP 598

                        241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGEREGPAGSPGFEGLPGAPGPPGEAGKPGEE 300
                            |||:||||||||||||:|||||||||||| |:|||||||| || |||||||||||||:
                        599 GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEGPAGSPGFQGLPGAPGPPGEAGKPGEQ 658

                        301 GVPGDLGAPGPGSGARGEPGFPGERGVEGPPGPPGPGADGAPGDDGADGAPGAPGSE 360
                            ||||||||||||| ||||| ||||||||||||||||||||||| ||||||||||||:
                        659 GVPGDLGAPGPGSGARGERGFPGERGVQGPPGPPGPRGANGAPGCNDGAKGDAGAPGAPGSQ 718

                        361 GAPGLEGMPGERGAAGLPGKGDRDAGPKGKADGSPGKDGVRGLTGPIGPGPPGAPGDK 420
                            |||:| ||||||||||:|||||||||:||||||||||||||||||| |||||||||||
                        719 GAPGLQGMPGERGAAGLPGEKGRDAGPKGKADGSPGKDGVRGLTGPIGPGPPGAPGDK 778

                        421 GBSGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPGADGEPGAKGEPGDAGKPGDAGPP 480
                            |||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
                        779 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPGADGQPGAKGEPGDAGKPGDAGPP 838

                        481 GPAGPAGPGPIGDVGAPGAKGARGSAGPPGPGATGFPGAAGRVGPPGPSGDAGPPGPPGPA 540
                            ||||||||||||| |:||:|||||||||||| |||||||||||||||||||| |||||||||
                        839 GPAGPAGPGPIGNVGAPGAKGARGSAGPPGPGATGFPGAAGRVGPPGPSGNAGPPGPPGPA 898

                        541 GKEG 544
                            ||||
                        899 GKEG 902
```

```
RESULT 15
ABG93947
ID   ABG93947 standard; protein; 1461 AA.
XX
AC   ABG93947;
XX
DT   26-NOV-2002  (first entry)
XX
DE   Human polypeptide orthologous to DACC-11.
XX
KW   Human; deer; rat; mouse; DACC; deer antler cartilage cell;
KW   cell stimulation; cell inhibition; cell growth; cell division;
KW   mesenchymal cell; chondrocyte; chondrogenesis; osteogenesis; growth;
KW   repair; regeneration; restoration; extracellular matrix;
KW   cartilaginous matrix; cartilage; disc; connective tissue; agonist;
KW   antagonist; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200264625-A1.
XX
PD   22-AUG-2002.
XX
PF   15-FEB-2002; 2002WO-AU000163.
XX
PR   15-FEB-2001; 2001AU-00003116.
XX
PA   (ADPP-) ADP PHARM PTY LTD.
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Roubin R,  Ghosh P;
XX
DR   WPI; 2002-643456/69.
XX
PT   stimulating or inhibiting cell growth and/or division, useful for
PT   stimulating chondrogenesis, cartilage, disc or connective tissue growth,
PT   repair, and/or regeneration, comprises administering deer antler
PT   cartilage gene.
XX
PS   Claim 13; Page 154-160; 214pp; English.
XX
CC   The invention discloses a method for stimulating or inhibiting cell
CC   growth and/or division which comprises contacting or inserting into an
CC   animal cell a polypeptide comprising one of the deer antler cartilage
```

```
CC   cell (DACC) clones disclosed. More particularly, the method relates to
CC   these polypeptides stimulating mesenchymal cell growth and/or division
CC   and to transfecting these cells and chondrocytes with vectors carrying
CC   the genes of these polypeptides capable of stimulating chondrogenesis,
CC   osteogenesis, growth, repair, regeneration and/or restoration of the
CC   extracellular matrix. The chondrocytes selectively express genes required
CC   to form a cartilaginous matrix. The DACC polypeptides and polynucleotides
CC   are useful for identifying an agent that modulates the activity of the
CC   polypeptide, for stimulating mesenchymal cell growth and/or division by
CC   exposing animal mesenchymal cells to conditioned media or its active
CC   fraction, obtained from deer antler cartilage cells, for inhibiting cell
CC   growth and/or division by inserting into an animal cell, a compound which
CC   inhibits the translation of the polynucleotide encoding the DACC. The
CC   method and the polypeptides are useful for stimulating mesenchymal cell
CC   growth and/or division or for stimulating chondrogenesis, cartilage, disc
CC   or connective tissue growth, repair, regeneration and/or restoration in
CC   an animal. The polynucleotides, polypeptides, agonists and antagonists
CC   may be used in treatment modalities, specifically in gene therapy. The
CC   polypeptides can be used as bait proteins in a two- or three-hybrid assay
CC   to identify other proteins, which bind to or interact with the
CC   polypeptide and are involved in modulating cell growth and/or division.
CC   The sequences presented in ABG93923-ABG93948 are the proteins encoded by
CC   the DACC cDNA clones
XX
SQ   Sequence 1461 AA;

Query Match             96.6%;   Score 2966;  DB 5;    Length 1461;
Best Local Similarity   95.6%;   Pred. No. 3.5e-169;
Matches  520;  Conservative  21;  Mismatches   3;   Indels   0;   Gaps   0;

Qy      1  GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAXGADGAPGIAGAPGFPGARGPSGPE  60
           |||||||||||||||||||||||||||||||:|||: ||||||||||||||||||||:
Db    359  GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ  418

Qy     61  GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEEGXPGARGPEPGPTGLP  120
           |||||||||||||:||| ||||||||||||||||::||||||||||:||||||||||||||
Db    419  GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEEGKRGARGPEPGPTGLP  478

Qy    121  GPPGERGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT  180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    479  GPPGERGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT  538

Qy    181  GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGARGEAGVMGPPGPKGAAGEPGKAGERGVP  240
           |||||||||||||||||||||||||||||||:|||::|||||||||||||||||||||||
Db    539  GSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGPPGPKGAAGEPGKAGERGVP  598

Qy    241  GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEGPAGSPGFPGLPGPAGPPGEAGKPGEE  300
           ||||||||||||||||||||||||||||||||||:||:|||||||||||||||||||::
Db    599  GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPPGPAGSPGFQGLPGPAGPPGEAGKPGEQ  658

Qy    301  GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGADGAPGDDGAKGDPAGAPGSE  360
           ||||||||||||||||||||||||:||:|||||||||||||||:|||||||||||:
Db    659  GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDPAGAPGSQ  718

Qy    361  GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  420
           |||||:||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db    719  GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  778

Qy    421  GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPGATGFPGAAGRVGPPGPSGDAGPP  480
           |||||||||||||||||||||:||||||||||||||||:|||||||||||||||||||||
Db    779  GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPGDGQPGAKGEPGDAGAKGDAGPP  838

Qy    481  GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPPGPPGPA  540
           |||||||||||::|||:|||||||||||||||||||||||||||||||||||||||||||
Db    839  GPAGPAGPPGPIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPA  898

Qy    541  GKEG  544
           ||||
Db    899  GKEG  902
```

Search completed: May 7, 2004, 16:28:20
Job time : 62 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      May  7, 2004, 16:27:10 ; Search time 23 Seconds
                                    (without alignments)
                                    1221.066 Million cell updates/sec

Title:       US-10-658-989A-4
Perfect score: 3070
Sequence:    1 GSEGPEGVRGEPGPGPGPAGA.........PGPSGDAGPPGPPGPAGKEG 544

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:    389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters:      389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    Issued Patents AA:*
         1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
         2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
         3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
         4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
         5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
         6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
          score greater than or equal to the score of the result being printed,
          and is derived by analysis of the total score distribution.

                            SUMMARIES
```

Result No.	Score	% Query Match	Match Length	DB	ID		Description
1	2966	96.6	1461	4	US-09-585-887-9	1885 61.4 1366 4 US-09-289-578-10	Sequence 9, Appli
2	2966	96.6	1461	4	US-09-289-578-9	1871 60.9 1078 3 US-08-963-825-21	Sequence 9, Appli
3	2966	96.6	1464	4	US-09-331-347C-21	1871 60.9 1078 4 US-09-500-811-21	Sequence 21, Appl
4	2857	93.1	822	3	US-09-219-849-49	1871 60.9 1078 4 US-09-570-573-21	Sequence 49, Appli
5	2842	92.6	1057	3	US-08-931-820-1	1871 60.9 1078 4 US-09-548-608-21	Sequence 1, Appli
6	2791.5	90.9	1341	4	US-08-963-825-18	1640.5 53.4 1064 1 US-08-642-255-62	Sequence 18, Appl
7	2791.5	90.9	1341	4	US-09-500-811-18	1600 52.1 720 3 US-09-219-849-4	Sequence 18, Appl
8	2791.5	90.9	1341	4	US-09-570-573-18	1600 52.1 777 3 US-08-642-255-53	Sequence 18, Appl
9	2791.5	90.9	1341	4	US-09-548-608-18	1522.5 49.6 492 4 US-08-468-996-12	Sequence 10, Appl
10	2226	72.5	1017	4	US-08-468-996-10	1502.5 48.9 1065 1 US-08-642-255-72	Sequence 10, Appl
11	2226	72.5	1060	3	US-08-931-820-3	1496.5 48.7 561 1 US-08-642-255-52	Sequence 3, Appli
12	2202	71.7	1418	3	US-08-963-825-20	1490.5 48.6 504 1 US-09-219-849-3	Sequence 20, Appl
13	2202	71.7	822	3	US-09-010-999-1	1447 47.1 633 1 US-08-642-255-73	Sequence 1, Appli
14	2202	71.7	1418	4	US-09-500-811-20	1414.5 46.1 960 3 US-09-219-849-5	Sequence 20, Appl
15	2202	71.7	1418	4	US-09-570-573-20	1406.5 45.8 829 1 US-08-642-255-132	Sequence 20, Appl
16	2202	71.7	1418	4	US-09-548-608-20	1406.5 45.8 829 1 US-08-397-633A-53	Sequence 20, Appl
17	2192	71.4	1442	2	US-08-316-650-12	1406.5 45.8 837 1 US-08-175-155-68	Sequence 12, Appl
18	2192	71.4	1442	5	PCT-US95-02251-12	1406.5 45.8 837 1 US-08-477-509B-103	Sequence 12, Appl
19	2156	70.2	595	3	US-09-219-849-48		Sequence 48, Appl
20	2156	70.2	595	3	US-09-219-849-50		Sequence 50, Appl
21	1901	61.9	1057	3	US-08-931-820-4		Sequence 4, Appli
22	1886	61.4	1024	3	US-08-931-820-2		Sequence 2, Appli
23	1886	61.4	1366	3	US-08-963-825-19		Sequence 19, Appl
24	1886	61.4	1366	4	US-09-500-811-19		Sequence 19, Appl
25	1886	61.4	1366	4	US-09-570-573-19		Sequence 19, Appl
26	1886	61.4	1366	4	US-09-548-608-19		Sequence 19, Appl
27	1885	61.4	1366	4	US-09-585-887-10		Sequence 10, App

```
                            ALIGNMENTS

RESULT 1
US-09-585-887-9
; Sequence 9, Application US/09585887
; Patent No. 6413742
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 22500230400
; CURRENT APPLICATION NUMBER: US/09/585,887
; CURRENT FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/289,578
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SEQ ID NO 9
;   LENGTH: 1461
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-585-887-9

Query Match             96.6%;  Score 2966;  DB 4;  Length 1461;
Best Local Similarity   95.6%;  Pred. No. 8.9e-177;
Matches  520;  Conservative  21;  Mismatches  3;  Indels  0;  Gaps  0;

Qy      1 GSEGPEGVRGEPGPGPGPAGAAGAGPGPGADGEPGAKGADGAPGFPGARGPSGPE 60
            |||||||||||||||||||||::||||:|||||| |||:::||:|||||||||
Db    359 GSEGPQGVRGEPGPGPGPAGAAGPAGNPGAGNPGAGNAPGIAGAPGFPGARGPSGPQ 418

Qy     61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPVGQGPPGVGEGKPGAGREGPGPTGLP 120
            |||||||||||||||||||||||||||:|||::||:||||||||||::||||:|||||||||
Db    419 GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPGAGEGKRGARGEPGPTGLP 478

Qy    121 GPPGERGGPGPGRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
            |||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db    479 GPPGERGGPGPGRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 538

Qy    181 GSPGSPGPDGKTGPGPAGEDGRPGPPGPPGGAAGEPGKAAGERGVP 240
            ||:||||||:||||:||||||:|||||||||||||||||||||||
Db    539 GSPGSPGDGKTGPPGPAGQDGRPGPGPPGGAARGQAVMGFPGPGKAAGERGVP 598

Qy    241 GPPGAVGPAGKDGEAGAAEGPGPAGPAGEREGEAGPSGPFBGLPGPGAPPGPEAGKPGEE 300
            |||::|||:||:|:|:||||:||||:|::|||:|||::|:||:|||:||:
```

```
Db  599 GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGPSPGFQGLPGPAGPPGEAGKPGEQ 658
Qy  301 GVPGDLGAPGPSGARGEPGPFGERGVEGPPGPAGPPGADGAPGDDGAKGDAGAPGAPGSE 360
        ||||| ||||||||||||:|| |||||:|||||||||||||||| || ||||||||||||:
Db  659 GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPPGADGAPGNDGAKGDAGAPGAPGSQ 718
Qy  361 GAPGLEGMPGERGAAGLPGPKGRDGDAGPKGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGPGDK 420
        |||||:|||||||||||||||| :|||||||||||||||||||||||||||||||||||||||||
Db  719 GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGPGDK 778
Qy  421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGFAGPPGADGEPGAKGEPGDAGAKGDAGPP 480
        |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db  779 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPP 838
Qy  481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPGPSGDAGPPGPGPA 540
        ||||||||||||:|:|||||||||||||||||||||||||||||||||||:|:||||||||
Db  839 GPAGPAGPPGPIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPGPSGNAGPPGPGPA 898
Qy  541 GKEG 544
        ||||
Db  899 GKEG 902
```

RESULT 2
US-09-289-578-9
```
; Sequence 9, Application US/09289578
; Patent No. 6428978
; GENERAL INFORMATION:
;   APPLICANT:  Olsen, David R
;   APPLICANT:  Chang, Robert
;   APPLICANT:  McMullin, Hugh
;   APPLICANT:  Hitzeman, Ronald A.
;   APPLICANT:  Chisholm, George
;   TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
;   TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
;   TITLE OF INVENTION: CELLS
;   FILE REFERENCE: 225002030400
;   CURRENT APPLICATION NUMBER: US/09/289,578
;   CURRENT FILING DATE: 1999-04-10
;   PRIOR APPLICATION NUMBER: 60/084,828
;   PRIOR FILING DATE: 1998-05-08
;   NUMBER OF SEQ ID NOS: 11
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 1461
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-289-578-9
```

```
Query Match          96.6%;  Score 2966;  DB 4;  Length 1461;
Best Local Similarity 95.6%;  Pred. No. 8.9e-177;
Matches 520;  Conservative 21;  Mismatches 3;  Indels 0;  Gaps 0;

Qy    1 GSEGPEGVRGEPGPPGPAGAAGPDGPGDGEPGAKGADGAPGIAGAPGFPGARGPSGPE 60
        |||||:||||||||||||||||||:||| || ||||:|||||||||||||||||||||:
Db  359 GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGPIAGAPGFPGARGPSGPQ 418
Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEEGKPGARGEPGPTGLP 120
        |||||||||::|||||||||||||||||||||||||:||||||||||:|||||||||||||
Db  419 GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEEGKRGARGEPGPTGLP 478
Qy  121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  479 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 538
Qy  181 GSPGSPGEDGKTGPPGPAGEDGRPGPPGPAGERGEAGVMGFPGPKGAAGEPGKAGERGVP 240
        |||||||:|||||||||||:|||||||||||||||:|||||||||||||||||||||||||
Db  539 GSPGSPGFDGKTGPPGPAGQDGRPGPPGPAGERGQAGVMGFPGPKGAAGEPGKAGERGVP 598
Qy  241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEEGPAGPSPGFEGLPGPAGPPGEAGKPGEE 300
        |||||||||||||||||:||||||||||||||:||||||||||:||||||||||||||||:
Db  599 GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGPSPGFQGLPGPAGPPGEAGKPGEQ 658
```

```
Qy  301 GVPGDLGAPGPSGARGEPGPFGERGVEGPPGPAGPPGADGAPGDDGAKGDAGAPGAPGSE 360
        ||||||||||||||||:|| |||||:|||||||||||||||| || ||||||||||||:
Db  659 GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPPGADGAPGNDGAKGDAGAPGAPGSQ 718
Qy  361 GAPGLEGMPGERGAAGLPGPKGRDGDAGPKGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGPGDK 420
        |||||:|||||||||||||||| :|||||||||||||||||||||||||||||||||||||||||
Db  719 GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGPGDK 778
Qy  421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGFAGPPGADGEPGAKGEPGDAGAKGDAGPP 480
        |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db  779 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPP 838
Qy  481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPGPSGDAGPPGPGPA 540
        ||||||||||||:|:|||||||||||||||||||||||||||||||||:|:||||||||
Db  839 GPAGPAGPPGPIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPGPSGNAGPPGPGPA 898
Qy  541 GKEG 544
        ||||
Db  899 GKEG 902
```

RESULT 3
US-09-331-347C-21
```
; Sequence 21, Application US/09331347C
; Patent No. 6617431
; GENERAL INFORMATION:
;   APPLICANT:  Meristem Therapeutics, S.A.
;   APPLICANT:  Meristem Therapeutics, S.A.
;   TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants,
;   TITLE OF INVENTION: obtaining Such and Their Uses
;   FILE REFERENCE: 1149-3
;   CURRENT APPLICATION NUMBER: US/09/331,347C
;   CURRENT FILING DATE: 1999-08-17
;   NUMBER OF SEQ ID NOS: 22
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
;   LENGTH: 1464
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-331-347C-21
```

```
Query Match          96.6%;  Score 2966;  DB 4;  Length 1464;
Best Local Similarity 95.6%;  Pred. No. 8.9e-177;
Matches 520;  Conservative 21;  Mismatches 3;  Indels 0;  Gaps 0;

Qy    1 GSEGPEGVRGEPGPPGPAGAAGPDGPGDGEPGAKGADGAPGIAGAPGFPGARGPSGPE 60
        |||||:||||||||||||||||||:||| || ||||:|||||||||||||||||||||:
Db  362 GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGPIAGAPGFPGARGPSGPQ 421
Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEEGKPGARGEPGPTGLP 120
        |||||||||::|||||||||||||||||||||||||:||||||||||:|||||||||||||
Db  422 GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEEGKRGARGEPGPTGLP 481
Qy  121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  482 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 541
Qy  181 GSPGSPGDGKTGPPGPAGEDGRPGPPGPAGERGEAGVMGFPGPKGAAGEPGKAGERGVP 240
        |||||||:|||||||||||:|||||||||||||||:|||||||||||||||||||||||||
Db  542 GSPGSPGFDGKTGPPGPAGQDGRPGPPGPAGERGQAGVMGFPGPKGAAGEPGKAGERGVP 601
Qy  241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEEGPAGPSPGFEGLPGPAGPPGEAGKPGEE 300
        |||||||||||||||||:||||||||||||||:||||||||||:||||||||||||||||:
Db  602 GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGPSPGFQGLPGPAGPPGEAGKPGEQ 661
Qy  301 GVPGDLGAPGPSGARGEPGPFGERGVEGPPGPAGPPGADGAPGDDGAKGDAGAPGAPGSE 360
        ||||||||||||||||:|| |||||:|||||||||||||||| || ||||||||||||:
Db  662 GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPPGANGAPGNDGAKGDAGAPGAPGSQ 721
Qy  361 GAPGLEGMPGERGAAGLPGPKGRDGDAGPKGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGPGDK 420
        |||||:|||||||||||||||| :|||||||||||||||||||||||||||||||||||||||||
Db  722 GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGPGDK 781
```

RESULT 4

US-09-219-849-49

```
Query Match          93.1%;  Score 2857;  DB 3;  Length 822;
Best Local Similarity  91.4%;  Pred. No. 3.1e-170;
Matches  497;  Conservative  28;  Mismatches  19;  Indels  0;  Gaps  0;

Qy    1 GSEGPEGVRGEPGPPGPAGAAGDPAGDGEPGAKGADGAPGIAGAPGFPGARGPSGPE    60
          |||||| |||||||||||||||  ||| || ||||||||||  |||||||||||||||||:
Db  183 GSEGPQGVRGEPGPPGPAGNPGADGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ  242

Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEEGKRGARGEPGPTGLP   120
          ||.|||||||.||||||||.||||||||||||||||||:||||||||||||||||||||||
Db  243 GPSGPPGPKGNSGEPGAPGNKGDTGAKGEPGPVGVEGPPGPAGEGKRGARGEPGPSGLP    302

Qy  121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGEAGRPGEAGLPGAKGLT          180
          ||||||||||||||||||||||||||||:||| ||||||||||||||||||
Db  303 GPPGERGGPGSRGFPGADGVAGPKGPSGERGAPGAPGPKGPSGEAGRPGEAGLPGAKGLT  362

Qy  181 GSPGSPGPGDGKTGPPGAGEDGRPGPPGPAGERGEEGPPGKGAAGEPGKAGERGVP     240
          ||||||||||||||||||| ||||: |||||||||||:||||||||||||||||||
Db  363 GSPGSPGPGDGKTGPPGAGQDGRPGAGQDGRPGARGQAGVMGFPGKGTAGEPGKAGERGLP  422

Qy  241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEEGPAGSPGFEGLPGPAGPGEAGKPGEE   300
          |||||||||||||||||  |||||||||||||| |||||||||:||||||||||||||||
Db  423 GPPGAVGPAGKDGEAGAQGPAGPAGERGEQGPAGSPGFQGLPGPAGRGQAGVMGFPGKGAAGERGVP  482

Qy  301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGPAGDGAPGDGAKGDAGAPGADGSA   360
          |||||||||||||||||.|| ||||| |||||||||:|||||||||||||||||||||||:
Db  483 GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGRNGAPGNDGAKGDTGAPGAPGSQ  542

Qy  361 GAPGLEGMPGERGAAGLPGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK          420
          |||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy  421 GESGPSGPAGPTGARGAPGDRGRGEPGPPGPAGFAGPPGADGEPGAKGEPGDAGAKGDAGPP  480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  782 GESGPSGPAGPTGARGAPGDRGRGEPGPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPP  841

Qy  481 GPAGPAGPPGPIGDVGPAGKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPPGPPGPA   540
          |||||||||||  ||||||||||||||||||||||| |||| ||||||||||||||||||:
Db  842 GPAGPAGPPGPIGNVGAPGAKGAKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPA  901

Qy  541 GKEG                                                        544
          ||||
Db  902 GKEG                                                        905
```

RESULT 5

US-08-931-820-1

```
Query Match          92.6%;  Score 2842;  DB 3;  Length 1057;
Best Local Similarity  91.9%;  Pred. No. 3.3e-169;
Matches  500;  Conservative  24;  Mismatches  20;  Indels  0;  Gaps  0;

Qy    1 GSEGPEGVRGEPGPPGPAGAAGDPAGDGEPGAKGADGAPGIAGAPGFPGARGPSGPE    60
          |||||| ||||||||||||| |  ||| || || ||||||||||||||||||||||||:
Db  201 GSEGPQGVRGEPGPPGPAGNPGADGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ  260

Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEEGKRGARGEPGPTGLP   120
          ||.|||||||.||||||||.|||||||||| ||:|: |||| |||||||||||||||||||
Db  261 GPGGPTGARGLVGEPGPAGSKGESGNKGEPGSAGPQGPPGPSGEEGKRGPNGEAGSAGPP    320

Qy  121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGEAGRPGEAGLPGAKGLT          180
          ||||||| ||||||||||| |||||||||||||||||||||||||||||||
Db  321 GPPGLRGGPGSRGFPGADGVAGPKGPAGERGSPGAGVPKGGSPGAGVPKGGSPGEAGRPGEAGLPGAKGLT  380

Qy  181 GSPGSPGPGDGKTGPPGAGEDGRPGPPGPAGERGEEGPPGKGAAGEPGKAGERGVP     240
          ||||||||||||||||||| ||||| |||||||||||||||||||||||||||||||
Db  381 GSPGSPGPDGKTGPPGPAGQDGRPGPPGPAGERGQAGVMGFPGKGAAGEPGKAGERGVP  440

Qy  241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEEGPAGSPGFEGLPGPAGPGEAGKPGEE   300
          |||||||||||||||||  ||||||||||||| |||||||| ||||||||||||||||||
Db  441 GPPGAVGPAGKDGEAGAQGPPGPAGERGBQGPPGPAGERGBQGPAGSPGFQGLPGPAGPGEAGKPGEQ  500
```

```
Qy    301  GVPGDLGAPGPSGARGEPGPPGERGVEGPPGCPAGPPGADGAPGDDGAKGDAGARGCAPGSE  360
           ||||||||||||||||||: ||||||:|  ||:||:||||||||||||:|||||| ||||:
Db    501  GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQ    560

Qy    361  GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK   420
           ||||:|||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db    561  GAPGLQGMPGERGAAGLPGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK            620

Qy    421  GESGPSGPAGPTGARGAPGDRGBEPGPPGPAGFAGPPGDAGEPGAKGEPGDAGAKGDAGPP  480
           ||||||||||||||||||||||:|||||||||||||||| :|||||||||||||||||||
Db    621  GESGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGDADGQPGAKGEPGDAGAKGDAGPP   680

Qy    481  GPAGPAGPPGIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPPGPPGPA    540
           |||||||||||:||||||||||||||||||||||||||||||||||||:||||||||||
Db    681  GPAGPAGPPGIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPA    740

Qy    541  GKEG  544
           ||||
Db    741  GKEG  744

RESULT 6
US-08-963-825-18
; Sequence 18, Application US/08963825
; Patent No. 6110689
; GENERAL INFORMATION:
;   APPLICANT: Qvist, Per
;   APPLICANT: Bonde, Martin
;   TITLE OF INVENTION: A Method for Assaying Collagen Fragments
;   TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
;   TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
;   TITLE OF INVENTION: Disorders Associated with the Metabolism of
;   NUMBER OF SEQUENCES: 21
;   CORRESPONDENCE ADDRESS:
;    ADDRESSEE: Darby & Darby PC
;    STREET: 805 Third Avenue
;    CITY: New York
;    STATE: New York
;    COUNTRY: USA
;    ZIP: 10022
;   COMPUTER READABLE FORM:
;    MEDIUM TYPE: Floppy disk
;    COMPUTER: IBM PC compatible
;    OPERATING SYSTEM: PC-DOS/MS-DOS
;    SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;    APPLICATION NUMBER: US/08/963,825
;    FILING DATE:
;    CLASSIFICATION: 436
;   PRIOR APPLICATION DATA:
;    APPLICATION NUMBER: US/08/187,319
;    FILING DATE: 21-JAN-1994
;   ATTORNEY/AGENT INFORMATION:
;    NAME: Gogoris, Adda C
;    REGISTRATION NUMBER: 29,714
;    REFERENCE/DOCKET NUMBER: 4305/08701
;   TELECOMMUNICATION INFORMATION:
;    TELEPHONE: 212-527-7700
;    TELEFAX: 212-753-6237
;    TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
;   SEQUENCE CHARACTERISTICS:
;    LENGTH: 1341 amino acids
;    TYPE: amino acid
;    TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   ORIGINAL SOURCE:
;    ORGANISM: Homo sapiens
;   IMMEDIATE SOURCE:
;    CLONE: COLLAGEN ALPHA 1 (I)
US-08-963-825-18

Query Match    90.9%;  Score 2791.5;  DB 3;  Length 1341;
```

```
Best Local Similarity  87.6%;  Pred. No. 5.5e-165;
Matches 500;  Conservative 21;  Mismatches 23;  Indels 27;  Gaps  4;

Qy      1  GSSGPEGVRGEPGPPGPPAGAAGPAGDPGADGE---PGAKGADGAPGIAGAPGPGAR---   54
           |:||||||:|||||||||||| : |:|||:|||   ||||||||||||||||||||||
Db    211  GASGPMGRGPGPPGPPGKBGBBGZAGKPGRPGZRGRGPPPZGARGLPGTAGLPGMKGHRGFS  270

Qy     55  --------GPSGPEG---------PG--GPPGPKGDSGEPGAPGSKGDTGAKGEPGPV     93
                   |:||:|          |:  |||||||||||||||||||||||||||||||
Db    271  GLBGAKGBGPAGPKGZGPGSPGZBGAPZMGPGPKGNSGEPGAPGSKGDTGAKGEPGPV    330

Qy     94  GVEGPPGPAGEEGKPGARGEPGPTGLPGPPGERGGPGSRGFPGADGVAGPKGPAGERGSP   153
           ||:||:||||||||:|||||||||||||||||||||||||||:||||||||||||||||
Db    331  GVQGP2GPAGEEGKRGARGEPGPTGLPGPPGERGGPGSRGFFGADGVAGPKGPAGERGSP   390

Qy    154  GPAGPKGSPGEAGRPGEAGLPGAKGLTGSPGSPGPDGKTGPPGAPGEDGRPGPPGPPGAR   213
           ||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||
Db    391  GPAGPKGSPGEAGRPGEAGLPGAKGLTGSPGSPGPDGKTGPPGAPGQDGRPGPPGPPGAR   450

Qy    214  GEAGVMGFPGPKGAAGEPGKAGERGVPGPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEE   273
           |:||||||||||||||||||||||||||||||||||||||||:|||||||||||||||:
Db    451  GQAGVMGFPGPKGAAGEPGKAGERGVPGPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQ   510

Qy    274  GPAGSPGFEGLPGPAGPPGEAGKPGEBGVPGDLGAPGPSGARGE2PGPFGERGVEGPEGPA  333
           |||||||:|||||||||||||||||:|||||||||||||||||:|||:||||||||||||
Db    511  GPAGSPGPQGLPGPAGPPGEAGKPGEQVPGDLGAPGPSGARGERGPPGERGVQGPPGPA   570

Qy    334  GPPGADGAPGDDGAKGDAGAPGAPGSEGAPGLEGMPGERGAAGLPGPKGDRGDAGPKGAD   393
           |||||:|||:|||||||||||||||:||||:||||||||||||||||||||||||||||
Db    571  GPRGANGAPGNDGAKGDAGAPGAPGSQGAPGLQGMPGERGAAGLPGPKGDRGDAGPKGAD   630

Qy    394  GSPGKDGVRGLTGPIGPPGPAGAPGDKGESGPSGPAGPTGARGAPGDRGBEPGPPGPAGFA  453
           ||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||
Db    631  GSPGKDGVRGLTGPIGPPGPAGAPGDKGESGPSGPAGPTGARGAPGDRGEPGPPGPAGFA   690

Qy    454  GPPGDAGEPGAKGEPGDAGAKGDAGPPGPAGPAGPPGIGDVGAPGAKGARGSAGPPGAT   513
           ||| :||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db    691  GPPGADGQPGAKGEPGDAGAKGDAGPPGPAGPAGPPGIGNVGAPGAKGARGSAGPPGAT   750

Qy    514  GFPGAAGRVGPPGPSGDAGPPGPPGPAGKEG  544
           |||||||||||||:|||||||||||||||||
Db    751  GFPGAAGRVGPPGSGNAGPPGPPGPAGKEG   781

RESULT 7
US-09-500-811-18
; Sequence 18, Application US/09500811
; Patent No. 6323314
; GENERAL INFORMATION:
;   APPLICANT: Qvist, Per
;   APPLICANT: Bonde, Martin
;   TITLE OF INVENTION: A Method for Assaying Collagen Fragments
;   TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
;   TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
;   TITLE OF INVENTION: Disorders Associated with the Metabolism of
;   NUMBER OF SEQUENCES: 21
;   CORRESPONDENCE ADDRESS:
;    ADDRESSEE: Darby & Darby PC
;    STREET: 805 Third Avenue
;    CITY: New York
;    STATE: New York
;    COUNTRY: USA
;    ZIP: 10022
;   COMPUTER READABLE FORM:
;    MEDIUM TYPE: Floppy disk
;    COMPUTER: IBM PC compatible
;    OPERATING SYSTEM: PC-DOS/MS-DOS
;    SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;    APPLICATION NUMBER: US/09/500,811
;    FILING DATE:
;    CLASSIFICATION:
;   PRIOR APPLICATION DATA:
```

APPLICANT: Bonde, Martin
TITLE OF INVENTION: A Method for Assaying Collagen Fragments
TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
TITLE OF INVENTION: Disorders Associated with the Metabolism of
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
  ADDRESSEE: Darby & Darby PC
  STREET: 805 Third Avenue
  CITY: New York
  STATE: New York
  COUNTRY: USA
  ZIP: 10022
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/570,573
  FILING DATE:
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 08/187,319
  FILING DATE:
ATTORNEY/AGENT INFORMATION:
  NAME: Gogoris, Adda C
  REGISTRATION NUMBER: 29,714
  REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 212-527-7700
  TELEFAX: 212-753-6237
  TELEX: 236687
INFORMATION FOR SEQ ID NO: 18:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 1341 amino acids
  TYPE: amino acid
  TOPOLOGY: linear
  MOLECULE TYPE: protein
  ORIGINAL SOURCE:
  ORGANISM: Homo sapiens
  IMMEDIATE SOURCE:
  CLONE: COLLAGEN ALPHA 1 (I)
US-09-570-573-18

Query Match   90.9%; Score 2791.5; DB 4; Length 1341;
Best Local Similarity 87.6%; Pred. No. 5.5e-166;
Matches 500; Conservative 21; Mismatches 23; Indels 27; Gaps 4;

Qy   1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGE---PGAKGADGAPGIAGAPGFPGAR---   54
          ||| | || |||| || | |||| || || ||||| || ||| ||| |||
Db   211 GASGPMGPRGPPGPPGPGKBGBBGZAGKPGRPGZRGPPGPZGARGLPGTAGLPGMKGHRGFS 270

Qy   55 --------GPSGPEG----------PG--GPPGPKGDSGEPGAPGSKGDTGAKGEPGPV   93
                  |: |:| || |||||||| |||||||||||||||||||||||
Db   271 GLBGAKGBAGPAGPKGZPGSPGZBGAPGZMGPPGPGKGNSGEPGAPGSKGDTGAKGEPGPV 330

Qy   94 GVEGPPGPAGEEGKPGARGEPGPTGLPGPPGERGGPGSRGFPGADGVAGPKGPAGERGSP 153
          |: |||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db   331 GVQGPPGPAGEEGKRGARGEPGPTGLPGPPGERGGPGSRGFPGADGVAGPKGPAGERGSP 390

Qy   154 GPAGPKGSPGEAGRPGEAGLPGAKGLTGSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGAR 213
          |||||||||||||||||||||||||||||||||||||||||||| |:||||||||||||
Db   391 GPAGPKGSPGEAGRPGEAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGAR 450

Qy   214 GEAGVMGFPGPKGAAGEPGKAGERGVPGPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEE 273
          |:||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||
Db   451 GQAGVMGFPGPKGAAGEPGKAGERGVPGPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQ 510

Qy   274 GPAGSPGFEGLPGPAGPPGPGAGKPGPEBGVPGDLGAPGPAGARGEPGFPGERGVEGPPGPA 333
          |||||||| ||||||||| |||| | |||||||||| |||| ||||||||:|||||||
Db   511 GPAGSPGFQGLPGPAGPPGEBAGKPGBEQGVPGDLGAPGPSGARGERGRFPGERGVQGPPGPA 570

Qy   334 GPPGADGAPGDDGAKGDAGAPGAPGSEGAPGLEGMPGERGAAGLPGPKGDRGDAGPKGAD 393
          || | |||||:||||||||||||||||:||||| |||||||||||||||||||||||||
Db   571 GPRGANGAPGNDGAKGDAGAPGAPGSQGAPGLQGMPGERGAAGLPGPKGDRGDAGPKGAD 630

Qy   394 GSPGKDGVRGLTGPIGPPGPAGAPGDKGESGSPSGPAGPTGARGAPGDRGRPGPPGPAGFA 453
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   631 GSPGKDGVRGLTGPIGPPGPAGAPGDKGESGSPSGPAGPTGARGAPGDRGRPGPPGPAGFA 690

Qy   454 GPPGADGEPGAKGEPGDAGAKGDAGPPGPAGPIGDVGAPGAKGARGSAGPPGAT   513
          |||||||:||||||||||||||||||||||||| ||||||||||||||||||
Db   691 GPPGADGQPGAKGEPGDAGAKGDAGPPGPAGPPGPAGPPGIGNVGAPGAKGARGSAGPPGAT 750

Qy   514 GFPGAAGRVGPPGPGPSGDAGPPGPPGPAGKEG   544
          ||||||||||||||||||| |||||||||||||
Db   751 GFPGAAGRVGPPGPGPSGNAGPPGPPGPAGKEG   781

RESULT 8
US-09-570-573-18
; Sequence 18, Application US/09570573
; Patent No. 6342361
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per

```
Db  211 GASGPMGPRGPPGPPGKBGBBGZAGKPCRPGPZGARGLPGTAGLPGMKGHRGFS     270
Qy   55 --------GPSGPEG---------PG--GPPGPKGDSGPPGAPGSKGDTGAKGEPGPV    93
Db  271 GLBGAKGBAGPAGPKG2PGSPGGZBGAP2MGPPGPKGNSGEPGAPGSKGDTGAKGEPGPV  330
Qy   94 GVBGPPGPAGEBGKPGARGEPGPTGLPGPPGERGGPGSRGPGADGVAGPKGPAGERGSP   153
Db  331 GVQGPPGPAGEBGKRGARGEPGPTGLPGPPGERGGPGSRGFPGADGVAGPKGPAGERGSP  390
Qy  154 GPAGPKGSPGEAGRPGEAGLPGAKGLTGSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGAR  213
Db  391 GPAGPKGSPGEAGRPGEAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGAR  450
Qy  214 GEAGVMGPFGPKGAAGEPGKAGERGVPGPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEE  273
Db  451 GQAGVMGPFGPKGAAGEPGKAGERGVPGPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQ  510
Qy  274 GPAGSPGFEGLPGPAGPPGPEAGKPGEEGVPGDLGAPGPSGARGEPGFGPGERGVEGPGPA 333
Db  511 GPAGSPGFQGLPGPAGPPGPEAGKPGEQGVPGDLGAPGPSGARGERGFPGERGVQGPGPA  570
Qy  334 GPPGADGAPGDDGAKGDAGAPAGPGSEGAPGLEGMPGERGAAGLPGPKGDRDAGPKGAD   393
Db  571 GPRGANGAPGNDGAKGDAGAPGASGQGAPGLQGMPGERGAAGLPGPKGDRGDAGPKGAD   630
Qy  394 GSPGKDGVRGLTGPIGPPGPAGAPGDKGESGSGPSGPAGPTGARGAPGDRGEPGPPGPAGFA 453
Db  631 GSPGKDGVRGLTGPIGPPGPAGAPGDKGESGSGPSGPAGPTGARGAPGDRGEPGPPGPAGFA 690
Qy  454 GPPGADGEPGAKGEPGDAGAKGDAGPPGPAGPAGPIGDVGAPGAKGARGSAGPPGAT    513
Db  691 GPPGADGQPGAKGEPGDAGAKGDAGPPGPAGPAGPIGNVGAPGAKGARGSAGPPGAT    750
Qy  514 GFPGAAGRVGPPGPSGDAGPPGPPGPAGKEG   544
Db  751 GFPGAAGRVGPPGPSGNAGPPGPPGPAGKEG   781
```

RESULT 10
US-08-468-996-10
; Sequence 10, Application US/08468996
; Patent No. 6645504
; GENERAL INFORMATION:
; APPLICANT: Weiner, Howard
; APPLICANT: Miller, Ariel
; APPLICANT: Zheng, Zhengi
; APPLICANT: Ahmad, Al-Sabbagh
; TITLE OF INVENTION: BYSTANDER SUPPRESSION OF TYPE 1 DIABETES BY ORAL ADMINISTRATION
; TITLE OF INVENTION: GLUCAGON
; FILE REFERENCE: 1010/16959-US3
; CURRENT APPLICATION NUMBER: US/08/468,996
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: US 07/843,752
; PRIOR FILING DATE: 1992-02-28
; PRIOR APPLICATION NUMBER: US 07/460,852
; PRIOR FILING DATE: 1990-02-21
; PRIOR APPLICATION NUMBER: US 07/596,936
; PRIOR FILING DATE: 1990-10-15
; PRIOR APPLICATION NUMBER: US 07/065,734
; PRIOR FILING DATE: 1987-06-24
; PRIOR APPLICATION NUMBER: US 07/454,486
; PRIOR FILING DATE: 1989-12-20
; PRIOR APPLICATION NUMBER: US 07/487,732
; PRIOR FILING DATE: 1990-03-02
; PRIOR APPLICATION NUMBER: US 07/551,632
; PRIOR FILING DATE: 1990-07-10
; PRIOR APPLICATION NUMBER: US 07/379,778
; PRIOR FILING DATE: 1989-07-14
; PRIOR APPLICATION NUMBER: US 07/607,826
; PRIOR FILING DATE: 1990-10-31
; PRIOR APPLICATION NUMBER: US 07/595,468

```
Qy  334 GPPGADGAPGDDGAKGDAGAPAGPCSEGAPGLEGMPGERGAAGLPGPKGCDRGDAGPKGAD  393
Db  571 GPRGANGAPGNDGAKGDAGAPGASGQGAPGLQGMPGERGAAGLPGPKGDRGDAGPKGAD   630
Qy  394 GSPGKDGVRGLTGPIGPPGPAGAPGDKGESGSGPSGPAGPTGARGAPGDRGEPGPPGPAGFA 453
Db  631 GSPGKDGVRGLTGPIGPPGPAGAPGDKGESGSGPSGPAGPTGARGAPGDRGEPGPPGPAGFA 690
Qy  454 GPPGADGEPGAKGEPGDAGAKGDAGPPGPAGPPGPIGDVGAPGAKGARGSAGPPGAT    513
Db  691 GPPGADGQPGAKGEPGDAGAKGDAGPPGPAGPPGPIGNVGAPGAKGARGSAGPPGAT    750
Qy  514 GFPGAAGRVGPPGPSGDAGPPGPPGPAGKEG   544
Db  751 GFPGAAGRVGPPGPSGNAGPPGPPGPAGKEG   781
```

RESULT 9
US-09-548-608-18
; Sequence 18, Application US/09548608
; Patent No. 6355442
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/548,608
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
US-09-548-608-18

Query Match   90.9%; Score 2791.5; DB 4; Length 1341;
Best Local Similarity   87.6%; Pred. No. 5.5e-166;
Matches 500; Conservative 21; Mismatches 23; Indels 27; Gaps 4;

```
Qy    1 GSEGPEGVRGEPGPPGPAGEAGPAGPGPGADGB---PGAKGADGAPGIAGAPGFPGAR---  54
```

; PRIOR FILING DATE: 1990-10-10
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
; LENGTH: 1017
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-468-996-10

Query Match   72.5%; Score 2226; DB 4; Length 1017;
Best Local Similarity 72.6%; Pred. No. 5.2e-131;
Matches 395; Conservative 39; Mismatches 110; Indels 0; Gaps 0;

```
Qy    1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGDADGAPGIAGAPGFPGARGPSGPE 60
          |||| :|:||||||||| :||:|| :||||||||:||||  ||||||||||||:|:
Db  184 GPEGAQGPRGEPGTPGSPGPAGASGNPGTDGIPGAKGSAGAPGIAGAPGFPGPRGPPDDPQ 243

Qy   61 GPGGPPGKDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEEGKPGARGEPGPTGLP 120
          ||||||||     |||  :|:| :|||||||||||||||
Db  244 GATGPLGPKGQTGKPGIAGFKGEQGPKGEPGPAGPQGPGPAGEGEKGKRGARGEPGGVGPI 303

Qy  121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
          ||||||| || ||||||  :|||||||||| :  |||| :|||| ||
Db  304 GPPGERGAPGNRGFPGQDGLAGPKGANGDPGRPGEPGLPGARGLT 363

Qy  181 GSPGSPGPDGKTGPPGPAGEDGRGPGPPGPAGEAGVMGFPGPKGAAGEPGKAGERGVP 240
          | |   ||| :|||||||||||| |:||:||| |  ||||| ||| |||:||
Db  364 GRPGGAGPQGKVGPGAPGEDGPGRPGPPGPGARGQPGVMGFPGPKGANGEPGKAGEKGLP 423

Qy  241 GPPGAVGPAGKDGEAGASGPPGPAGPAGERGEEGPAGSPGFEGLPGPAGPGEAGKPGSE 300
          |||| |||:| |:|||  ||||||  |||| ||  |||:| |||||||||||
Db  424 GAPGLRGLPGKDGETGAEGPPGPAGERGEQAGPGPSGFQGLLPGPGPPGEAGKPGDQ 483

Qy  301 GVPGDLGAPGPSGARGEPGPPGERGVEGPPGPAGPGADGAPGADGAPGAPGSE 360
          |:  ||||||:|||||:||||  || ||||||||   || :||| |:
Db  484 GVPGEAGAPGLVGPRGERGFPGERGSPGAQGLQGPRGLPGTPGTDGPKGASGPAGPPGAQ 543

Qy  361 GAPGLEGMPGERGAAGLPGPKGDGDAGKGEPGDAGAKGDAGPP 420
          |||| ||||| || || | :| || || :|||||:|| |||
Db  544 GPPGLQGMPGERGAAGIAGPKGDRGDVGEKGPGAPGDGGRCLTGPIGPPGPAGANGEK 603

Qy  421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGDADGPKGDAGAP 480
          ||||| ||| ||||||||:| ||||| |||  |||  | ||:| :|||
Db  604 GEVGPPGPAGSGARGAPGERGETGPPGPAGPAGPGDADGQPGAKGEQEAGQKGDAGAP 663

Qy  481 GPAGPAGPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPGSDAGPPGPGPA 540
          |||||| |||||||||||||||||||||||||||||| |||||||||||:||||||| ||
Db  664 GPQGPSGAPGPQGPTGVTGPKGARGAQGPPGATGFPGAAGRVGPPGSNGNPGPPGPGPS 723

Qy  541 GKEG 544
          ||:|
Db  724 GKDG 727
```

RESULT 11
US-08-931-820-3
; Sequence 3, Application US/08931820
; Patent No. 6010863
GENERAL INFORMATION:
;   APPLICANT:
;   TITLE OF INVENTION: Assay for collagen degradation
;   NUMBER OF SEQUENCES: 4
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/931,820
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: EP 96202596.1

; FILING DATE:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1060 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
;   ORGANISM: Homo sapiens
;   TISSUE TYPE: Collagen type II
US-08-931-820-3

Query Match   71.7%; Score 2202; DB 3; Length 1060;
Best Local Similarity 71.9%; Pred. No. 1.7e-129;
Matches 391; Conservative 40; Mismatches 113; Indels 0; Gaps 0;

```
Qy    1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGDADGAPGIAGAPGFPGARGPSGPE 60
          |||| :|:||||||||| :||:|| :||||||||:||||  ||||||||||||:|:
Db  203 GPEGAQGPRGEPGTPGSPGPAGASGNPGTDGIPGAKGSAGAPGIAGAPGFPGPRGPPDDPQ 262

Qy   61 GPGGPPGKDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEEGKPGARGEPGPTGLP 120
          ||||||||     |||  :|:| :|||||||||||||||
Db  263 GATGPLGPKGQTGKPGIAGFKGEQGPKGEPGPAGPQGPGPAGEGEKGKRGARGEPGGVGPI 322

Qy  121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
          ||||||| || ||||||  :|||||||||| :  |||| :|||| ||
Db  323 GPPGERGAPGNRGFPGQDGLAGPKGAPGERGPSGLAGPKGANGDPGRPGEPGLPGARGLT 382

Qy  181 GSPGSPGPDGKTGPPGPAGEDGRGPGPPGPAGEAGVMGFPGPKGAAGEPGKAGERGVP 240
          | |   ||| :|||||||||||| |:||:||| |  ||||| ||| |||:||
Db  383 GRPGGAGPQGKVGPGAPGEDGRPGPGPQGPGPAGEDGPGVWGFPGPKGANGEPGKAGEKGLP 442

Qy  241 GPPGAVGPAGKDGEAGASGPPGPAGPAGERGEEGPAGSPGFEGLPGPAGPGEAGKPGEE 300
          |||| |||:| |:|||  ||||||  |||| ||  |||:| |||||||||||
Db  443 GAPGLRGLPGKDGETGAEGPPGPAGERGEQAGPGPSGFQGLLPGPGPPGEQGKAGEKGLP 502

Qy  301 GVPGAVGPAGKDGEAGABGPPGPAG2AGEGPAGSPGFEGLPGPAGPPPGEAGKPGEE 360
          |:  ||||||:|||||:||||  || ||||||||   || :||| |:
Db  503 GVPGEAGAPGLVGPRGERGFPGERGSPGAQGLQGPRGLPGTPGTDGPKGASGPAGPPGAQ 562

Qy  361 GAPGLEGMPGERGAAGLPGPKGDGDAGKGEPGDAGAKGDAGPP 420
          |||| ||||| || || | :| || || :|||||:|| |||
Db  563 GPPGLQGMPGERGAAGIAGPKGDRGDVGEKGPGAPGDGGRCLTGPIGPPGPAGANGEK 622

Qy  421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGDADGPKGDAGAP 480
          ||||| ||| ||||||||:| ||||| |||  |||  | ||:| :|||
Db  623 GEVGPPGPAGSGARGAPGERGETGPPGPAGPAGPGDADGQPGAKGEQEAGQKGDAGAP 682

Qy  481 GPAGPAGPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPGSDAGPPGPGPA 540
          |||||| |||||||||||||||||||||||||||||| |||||||||||:||||||| ||
Db  683 GPQGPSGAPGPQGPTGVTGPKGARGAQGPPGATGFPGAAGRVGPPGSNGNPGPPGPGPS 742

Qy  541 GKEG 544
          ||:|
Db  743 GKDG 746
```

RESULT 12
US-08-963-825-20
; Sequence 20, Application US/08963825
; Patent No. 6110689
GENERAL INFORMATION:
;   APPLICANT: Qvist, Per
;   APPLICANT: Bonde, Martin
;   TITLE OF INVENTION: A Method for Assaying Collagen Fragments
;   TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
;   TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
;   TITLE OF INVENTION: Disorders Associated with the Metabolism of
;   NUMBER OF SEQUENCES: 21
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Darby & Darby PC

```
Qy   481  GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPPGAAGRVGPPGPSGDAGPPGPPGPA  540
               |:||:||:|| ||||||||||||||||||||||||||||||||||||:|:||||||
Db   795  GPQGPSGAPGPGPQGPTGVTGPKGARGAQGPPGATGFPGAAGRVGPPGSNGNPGPPGPPGPS  854

Qy   541  GKEG  544
               ||:|
Db   855  GKDG  858
```

RESULT 13
US-09-010-999-1
; Sequence 1, Application US/09010999
; Patent No. 6132976
; GENERAL INFORMATION:
;   APPLICANT: Poole, Anthony R.
;   APPLICANT: Hollander, Anthony P.
;   APPLICANT: Billinghurst, R. C.
;   TITLE OF INVENTION: IMMUNOASSAYS FOR THE MEASUREMENT OF
;   TITLE OF INVENTION: COLLAGEN DENATURATION AND CLEAVAGE IN CARTILAGE
;   NUMBER OF SEQUENCES: 16
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Foley & Lardner
;   STREET: 3000 K Street, N.W., Suite 500
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: USA
;   ZIP: 20007-5109
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC Compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/010,999
;   FILING DATE: 22-JAN-1998
;   CLASSIFICATION: 4335
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/448,501
;   FILING DATE: 17-JUL-1995
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/984,123
;   FILING DATE: 04-DEC-1992
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Bent, Stephen A.
;   REGISTRATION NUMBER: 29,768
;   REFERENCE/DOCKET NUMBER: 032931/0212
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202)672-5300
;   TELEFAX: (202)672-5399
;   TELEX: 904136
;   INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1418 amino acids
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   ORIGINAL SOURCE:
;   ORGANISM: Human Type II Collagen
US-09-010-999-1

Query Match   71.7%; Score 2202; DB 3; Length 1418;
Best Local Similarity 71.9%; Pred. No. 2.1e-129;
Matches 391; Conservative 40; Mismatches 113; Indels 0; Gaps 0;

Qy   1 GSEGPEGVRGEPGPPGPAGAAGPADPGDADGEPGAKGADGAPGIAGAPGFPGARGPSGPE   60
               |||||||||||||||||||||||:|||| ||||:||||||||| ||||||:||||:
Db   315 GPEGAQGPGRGEPGTPGSPGPAGAGASGNPGTDGIPGAKGSAGAPGIAGAPGFPGPRGPPDPQ 374

Qy   61 GPGGPPGPKGDGEPGAPGSKGDTGAKGEPGPGVGVEGPPGPAGEEGKPGARGEPGPTGLP 120
               |||| |||||||||||||:|||||||| |:||||||||||| ||||||||||||||
Db   375 GATGPLGPKGGQTGKPGIAGFKGEQGPKGEPGPAGPQGAPGAGEEGKRGARGEPGGVGPI 434
```

```
Qy   121  GPPGERGGPGSRGFPGADGVAGFKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT  180
               |||||:||||:|||||  ||||||||||||||||||:||||||:|:||||||||||||
Db   435  GPPGERGAPGNRGFPGQDGLAGPKGAPGERGPSGLAGPKGANGDPGRPGEPGLPGARGLT  494

Qy   181  GSPGSPGPGKTGPPGPAGEDGRPGPPGPGARGEAGVMGFPGPKGAAGEPGKAGERGVP  240
               ||||||||| ||||||||||||||||:|||||||||||||||||||||:|:||
Db   495  GRPGDAGPQGKVGPSGPAGPEDGRPGPGPGQGARGQPGVMGFPGPKGANGEPGKAGEKGLP  554

Qy   241  GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEEGPAGSPGPEGLPGPAGPPGEAGKPGEE  300
               |||:|||||||:||||||||||||:|||||||||||:|||||||:||||||||||||||
Db   555  GAPGLRGLPGKDGDGTGAEGPPGPAGPAGERGEQGAPGEGRGEQGAPGSGFQGLPGPPGPPGEGGKPGDQ  614

Qy   301  GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGADGAPGDDGAKGDAGAPGSE  360
               ||:|||||||||||||||  ||||||:|||||:|||||:|:|||||||||||:
Db   615  GVPGEAGAPGLVGPRGERGFPGERGSPGAQGLQGPRGLPGTPGTDGPKGASGPAGPPGAQ  674

Qy   361  GAPGLEGMPGERGAAGLPGPKDRDGAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  420
               ||:||||||||||||||:||||:|:||  ||||||||||:|:||||||||||||:
Db   675  GPPGLQGMPGERGAAGIAGPKGDRGDVGEKGPEGAPGKDGGRGLTGPIGPPGPAGANGEK  734

Qy   421  GESGPSGPAGPTGARGARGDRGERGPPGPPGPAGPAGPGDGEPGAKGEPGDAGAKGDAGPP  480
               ||||||:||||||:||||||||:|:|||||:||:|||||||||:|:|||:|||
Db   735  GEVGPPGPAGSAGARGAPGERGETGPPGPVTSGIAGPPGGAPGPAGEPGDGGVGPI  794
```

```
Qy     1  GSGGPEGVRGEPGPPGPAGAAGPADGEPGAKGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE  60
               |||:|||||||||||||||||||||||||||||||||||||||||||||||||||:
Db   315  GPEGAQGPGRGEPGTPGSPGPAGAGASGNPGTDGIPGAKGSAGAPGIAGAPGFPGPRGPPDPQ  374

Qy    61  GPGGPPGPKGDGEPGAPGSKGDTGAKGEPGPGVGVEGPPGPAGEEGKPGARGEPGPTGLP  120
               |||| |||||||||||||:||||||||  |:|||||||||||  ||||||||||||||
Db   375  GATGPLGPKGGQTGKPGIAGFKGEQGPKGEPGPAGPQGAPGAGEEGKRGARGEPGGVGPI  434
```

Query Match   71.7%; Score 2202; DB 3; Length 1418;
Best Local Similarity 71.9%; Pred. No. 2.1e-129; Indels 0; Gaps 0;
Matches 391; Conservative 40; Mismatches 113;

```
Qy     1  GSEGPEGVRGEPGPPGPAGAAGPADPGDADGEPGAKGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE  60
               ||:||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db   315  GPEGAQGPGRGEPGTPGSPGPAGAGASGNPGTDGIPGAKGSAGAPGIAGAPGFPGPRGPPDPQ  374

Qy    61  GPGGPPGPKGDGSGEPGAPGSKGDTGAKGEPGPGVVEGPPGPVGEGPPGPAGEGKPGARGEPGPTGLP  120
               |||| ||||||||||||:||||||||  |:|||||||||||  |||||||||||||||
Db   375  GATGPLGPKGGQTGKPGIAGFKGEQGPKGEPGPAGPQGAPGAGEESGKRGARGEPGGVGPI  434
```

US-08-963-825-20

; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/963,825
;   FILING DATE:
;   CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/08/187,319
;   FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
;   NAME: Gogoris, Adda C
;   REGISTRATION NUMBER: 29,714
;   REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 212-527-7700
;   TELEFAX: 212-753-6237
;   TELEX: 236687
; INFORMATION FOR SEQ ID NO: 20:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1418 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   ORIGINAL SOURCE:
;   ORGANISM: Homo sapiens
;   IMMEDIATE SOURCE:
;   CLONE: COLLAGEN -ALPHA 1 (II)
US-08-963-825-20

RESULT 14 (right column)

```
SEQUENCE CHARACTERISTICS:
  LENGTH: 1418 amino acids
  TYPE: amino acid
  TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
  ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
  CLONE: COLLAGEN -ALPHA 1 (II)
US-09-500-811-20

Query Match            71.7%;  Score 2202; DB 4; Length 1418;
Best Local Similarity  71.9%;  Pred. No. 2.1e-129;
Matches  391; Conservative  40; Mismatches  113; Indels  0; Gaps  0;

Qy    1 GSEGPEGVRGEPGPPGPAGAAGPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE   60
         |||:|:||||||||:::||||:||:::|||||||||||:|::|:
Db  315 GPPGAQGPRGEPGETPGSPGPAGASGNPGTDGIPGAKGSAGAPGIAGAPGFPGPPDPQ  374

Qy   61 GPGPPPKGDSGEPGAPGSKDTGAKGEPGPGVGVEGPPGPAGEEGKPGARGEPGPTGLP  120
         |||:||:||||::|||||||||:|||||||::|
Db  375 GATGPLGPKGQTGKPGIAGFKGQGKPGKEPGPAGPQGAPGPAGBEGKRCARGEPGGVGPI  434

Qy  121 GPPGERGRGPGSRGFPGADGVAGPGKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT  180
         ||||||||:|:||||||:||:||:|||||:|||:|:||:||||||||||
Db  435 GPPGERGAPGNRGFPGQDGLAGPKGANGDPGRPGEPGLPGARGLT  494

Qy  181 GSPGSPGPDGPKTGPPGPAGEDGRPGPPGPPGAGEAGVMGFPGPKGAAGEPGKAGERGVP  240
         |||:|:||||:||||||||||:|||||::|||||:|||||||||:|
Db  495 GRPGDAGPQGKVGPSGAPGEDGRPGPPGPGQGARGQPGVMGFPGPKGANGEPGKAGEKGLP  554

Qy  241 GPPGAVGPAGKDGEAGAEGPPGPAGERGEPGPAGSPGFEGLPGAGVMGFPGPKGEAGKPGEE  300
         :|||:||||||||||:||||||||:|::||:||||||:::|||||||:||||:|||||:
Db  555 GAPGLRGLPGKDGETGAEGPPGPAGPAGERGEQGARGERGQGAPGPSGFQGLPGPPGPPGEGGKPGDQ  614

Qy  301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPGDGAKGDAGAPGAPGSE  360
         |||:||:|||||:|:||||:||||:|||||:|
Db  615 GVPGEAGAPGLVGPRGERGFPGERGSPGAQGLQGPRGLPGTPGTDGPKGASGPAGPGAQ  674

Qy  361 GAPGLEGMPGERGAAGLPGKDGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  420
         |||||||||||||:||||:|:||||:|||||||||||||||:||||||
Db  675 GPPGLQGMPGERGAAGIAGPKGDRGDVGEKGPEGAPGKDGGRGLTGPIGPPGPAGANGEK  734

Qy  421 GESGPSGPAGPTGARGAPGAGERGETGPPGTSGIAGPPGADGQPGAKGEQGEAGQKGDAGAP  480
         |||||||:|||:|||||||||:|:|||:|||:|||||:|||||:|||
Db  735 GEVGPPGPAGSAGARGAPGERGETGPPGTSGIAGPPGADGQPGAKGEQGEAGQKGDAGAP  794

Qy  481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPGSGDAGPPGPGPA  540
         :||||||||:|:|||||||:||:|||::|:
Db  795 GPQGPSGAPGPQGPTGVTGPKGARGAQGPPGATGFPGAAGRVGPPGSNGNPGPGPGPS  854

Qy  541 GKEG  544
        |||:|
Db  855 GKDG  858
```

RESULT 14 (left column)

```
Qy  121 GPPGERGRGPGSRGFPGADGVAGPGKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT  180
         ||||||||:|:||||||:||:||:|||||:|||:|:||:||||||||||
Db  435 GPPGERGAPGNRGFPGQDGLAGPKGANGDPGRPGEPGLPGARGLT  494

Qy  181 GSPGSPGPDGPKTGPPGPAGEDGRPGPPGPPGAGEAGVMGFPGPKGAAGEPGKAGERGVP  240
         |||:|:||||:||||||||||:|||||::|||||:|||||||||:|
Db  495 GRPGDAGPQGKVGPSGAPGEDGRPGPPGPGQGARGQPGVMGFPGPKGANGEPGKAGEKGLP  554

Qy  241 GPPGAVGPAGKDGEAGAEGPPGPAGERGEPGPAGSPGFEGLPGAGVMGFPGPKGEAGKPGEE  300
         :|||:||||||||||:||||||||:|::||:||||||:::|||||||:||||:|||||:
Db  555 GAPGLRGLPGKDGETGAEGPPGPAGPAGERGEQGARGERGQGAPGPSGFQGLPGPPGPPGEGGKPGDQ  614

Qy  301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPGDGAKGDAGAPGAPGSE  360
         |||:||:|||||:|:||||:||||:|||||:|
Db  615 GVPGEAGAPGLVGPRGERGFPGERGSPGAQGLQGPRGLPGTPGTDGPKGASGPAGPGAQ  674

Qy  361 GAPGLEGMPGERGAAGLPGKDGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  420
         |||||||||||||:||||:|:||||:|||||||||||||||:||||||
Db  675 GPPGLQGMPGERGAAGIAGPKGDRGDVGEKGPEGAPGKDGGRGLTGPIGPPGPAGANGEK  734

Qy  421 GESGPSGPAGPTGARGAPGAGERGETGPPGTSGIAGPPGADGQPGAKGEQGEAGQKGDAGAP  480
         |||||||:|||:|||||||||:|:|||:|||:|||||:|||||:|||
Db  735 GEVGPPGPAGSAGARGAPGERGETGPPGTSGIAGPPGADGQPGAKGEQGEAGQKGDAGAP  794

Qy  481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPGSGDAGPPGPGPA  540
         :||||||||:|:|||||||:||:|||::|:
Db  795 GPQGPSGAPGPQGPTGVTGPKGARGAQGPPGATGFPGAAGRVGPPGSNGNPGPGPGPS  854

Qy  541 GKEG  544
        |||:|
Db  855 GKDG  858
```

```
RESULT 14
US-09-500-811-20
; Sequence 20, Application US/09500811
; Patent No. 6323314
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,811
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 20:
```

```
RESULT 15
US-09-570-573-20
; Sequence 20, Application US/09570573
; Patent No. 6342361
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
```

```
'COUNTRY: USA
 ZIP: 10022
COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/570,573
 FILING DATE:
 CLASSIFICATION:
PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/187,319
 FILING DATE:
ATTORNEY/AGENT INFORMATION:
 NAME: Gogoris, Adda C
 REGISTRATION NUMBER: 29,714
 REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-527-7700
 TELEFAX: 212-753-6237
 TELEX: 236687
INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1418 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: COLLAGEN -ALPHA 1 (II)
US-09-570-573-20
```

```
Query Match      71.7%;   Score 2202;   DB 4;   Length 1418;
Best Local Similarity   71.9%;   Pred. No. 2.1e-129;
Matches 391;   Conservative   40;   Mismatches 113;   Indels   0;   Gaps   0;

Qy      1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGPGADGEPGAKGADGAPGIAGAGPPGARGPSGPE 60
Db    315 GPEGAQGPRGEPGTPGSPGPAGASGNPGTDGIPGAKGSAGAPGIAGAGPPGPRGPPRGPPDPQ 374

Qy     61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGAGEEGKPGARGEPGPTGLP 120
Db    375 GATGPLGPKGQTGKPGIAGFKGEQGPKGEPGPAGPQGAPGPAGEEGKRGARGEPGGVGPI 434

Qy    121 GPPGERGGPGSRGFPGADGVAGPKGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
Db    435 GPPGERGAPGNRGFPGQDGLAGPKGAPGERGPSGLAGPKGANGDPGRPGEPGLPGARGLT 494

Qy    181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGARGEAGVMGFPGPKGAAGEPGKAGERGVP 240
Db    495 GRPGDAGPGQGKVGPSGAPGEDGRPGPPGPQGARGQPGVMGFPGPKGANGEPGKAGEKGLP 554

Qy    241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEEGPAGSPGEGLPGPAGPPGEAGKPGEE 300
Db    555 GAPGLRGLPGKDGETGAEGPPGPAGPAGERGEQGAPGPSGFQGLPGPPGPPGEGGKPGDQ 614

Qy    301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGADGAPGDDGAKGDAGAPGAPGSE 360
Db    615 GVPGEAGAPGLIVGPRGERGFPGERGSPGAQGLQGPRGLPGTPGTDGPKGASGPAGPPGAQ 674

Qy    361 GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLITGPIGPPGPAGAPGDK 420
Db    675 GPGLQGMPGERGEAAGIAGPKGDRGDVGEKGPEGAPGKDGRGLIGPIGPPGPAGANGEK 734

Qy    421 GESGPSGPAGPPTGARGAPGDRGEPGPPGPAGPAGPPGADGEPGAKGEPGDAGAKGDAGPP 480
Db    735 GEVGPPGPAGSAGARGAPGERGETGPPGTSGIIAGPPGADGQPGAKGEBQGEAGQKGDAGAP 794

Qy    481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPPGPPGPA 540
Db    795 GPQGPSGAPGPQGPTGVTGPKGARGAQGPPGATGFPGAAGRVGPPGSNGNPGPPGPGPS 854

Qy    541 GKEG 544
          ||:|
Db    855 GKDG 858
```

Search completed: May 7, 2004, 16:30:57
Job time : 25 secs

OM protein - protein search, using sw model

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore version 5.1.6

Run on: May 7, 2004, 16:28:25 ; Search time 49 Seconds
(without alignments)
3081.557 Million cell updates/sec

Title: US-10-658-989A-4
Perfect score: 3070
Sequence: 1 GSEGPEGVRGEPGPGPGAGA..........PGPSGDAGPGPGPPGPAGKEG 544

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	2966	96.6	1464	12	US-09-918-715-261	Sequence 261, App
2	2966	96.6	1464	14	US-10-060-036-159	Sequence 159, App
3	2966	96.6	1464	14	US-10-171-311-36	Sequence 36, Appl
4	2966	96.6	1464	14	US-10-216-705-21	Sequence 21, Appl
5	2966	96.6	1464	14	US-10-149-352-2	Sequence 2, Appli
6	2966	96.6	1464	14	US-10-177-293-65	Sequence 65, Appl
7	2966	96.6	1464	14	US-10-301-822-28	Sequence 28, Appl
8	2966	96.6	1464	14	US-10-291-265-243	Sequence 243, App
9	2936	95.6	1463	15	US-10-402-089-2	Sequence 2, Appli
10	2936	95.6	1463	15	US-10-402-072A-2	Sequence 2, Appli
11	2857	93.1	822	13	US-10-342-331-49	Sequence 49, Appl
12	2816.5	91.7	1449	15	US-10-402-089-8	Sequence 8, Appli
13	2816.5	91.7	1449	15	US-10-402-072A-8	Sequence 8, Appli
14	2791.5	90.9	1341	15	US-10-058-124-18	Sequence 18, Appl
15	2209	72.0	1014	14	US-10-194-441A-1	Sequence 1, Appli
16	2202	71.7	1418	14	US-10-058-124-20	Sequence 20, Appl
17	2189	71.3	1014	14	US-10-194-441A-48	Sequence 48, Appl
18	2156	70.2	595	15	US-10-342-331-48	Sequence 48, Appl
19	2156	70.2	595	15	US-10-342-331-50	Sequence 50, Appl
20	2045	66.6	500	12	US-10-232-175-22	Sequence 22, Appl
21	2045	66.6	662	12	US-10-232-175-33	Sequence 33, Appl
22	1924.5	62.7	1466	15	US-10-402-089-12	Sequence 12, Appl
23	1924.5	62.7	1466	15	US-10-402-072A-12	Sequence 12, Appl
24	1923	62.6	1366	15	US-10-402-089-10	Sequence 10, Appl
25	1923	62.6	1366	15	US-10-402-072A-10	Sequence 10, Appl
26	1921.5	62.6	1496	14	US-10-177-293-70	Sequence 70, Appl
27	1921.5	62.6	1496	15	US-10-301-822-35	Sequence 35, Appl
28	1921.5	62.6	1496	15	US-10-236-031B-74	Sequence 74, Appl
29	1906.5	62.1	1466	15	US-10-402-089-4	Sequence 4, Appli
30	1906.5	62.1	1466	15	US-10-402-089-6	Sequence 6, Appli
31	1906.5	62.1	1466	15	US-10-402-072A-4	Sequence 4, Appli
32	1906.5	62.1	1466	15	US-10-402-072A-6	Sequence 6, Appli
33	1901	61.9	1466	12	US-09-918-715-226	Sequence 226, App
34	1901	61.9	1466	12	US-10-257-021-72	Sequence 72, Appl
35	1901	61.9	1466	14	US-10-177-293-68	Sequence 68, Appl
36	1901	61.9	1466	15	US-10-301-822-33	Sequence 33, Appl
37	1886	61.4	1466	12	US-10-257-021-76	Sequence 76, Appl
38	1886	61.4	1366	14	US-10-058-124-19	Sequence 19, Appl
39	1885	61.4	1366	12	US-09-918-715-214	Sequence 214, App
40	1885	61.4	1366	12	US-09-918-715-234	Sequence 234, App
41	1885	61.4	1366	12	US-09-918-715-265	Sequence 265, App
42	1885	61.4	1366	14	US-10-171-311-38	Sequence 38, Appl
43	1885	61.4	1366	14	US-10-301-822-31	Sequence 31, Appl
44	1871	60.9	1078	14	US-10-058-124-21	Sequence 21, Appl
45	1788.5	58.3	714	14	US-10-233-885-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1
US-09-918-715-261
; Sequence 261, Application US/09918715
; Publication No. US2003001715A1
; GENERAL INFORMATION:
; APPLICANT: Brad St. Croix
; APPLICANT: Bert Vogelstein
; APPLICANT: Kenneth Kinzler
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00134
; CURRENT APPLICATION NUMBER: US/09/918,715
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 60/222,599
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 60/224,360
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 261
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-918-715-261

Query Match   96.6%; Score 2966; DB 12; Length 1464;
Best Local Similarity 95.6%; Pred. No. 5.4e-146;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

Qy   1 GSEGPEGVRGEPGPPGPAGAAGPAGPGDGADGPGPAGKGADGAPGIAGAPGFPGARGPSGPE 60
           ||||||||||||||||||||||||:||::|||||||:||||:||:|||:|||||||||||:
Db   362 GSEGPQGVRGEPGPPGPAGAAGPAGPAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ 421

Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPGVGVEGPPGPAGEEGKPGAPGEBGPTGLP 120
           |||||||||||:||||||||||||||||||||||||:||||||||||||||||||||||||
Db   422 GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPGVGVQGPPGPAGEEGKRGARGEPGPFGLP 481

```
Qy  121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGKGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  482 GPPGERGGPGSRGFPGADGVAGPKGPAGERGKGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 541

Qy  181 GSPGSPGPDGKTGPPGPAGEDGRPGPGPAGERGEGPGAGEPGKGAAGEPGKAGERGVP 240
        |||||||||||||||:|                  |       ::|||||||||||||
Db  542 GSPGSPGPDGKTGPPGPAGEDGRPGPGPPGARGQDGRPGPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGVP 601

Qy  241 GPPGAVGPAGKDGEAGAEGPPGPAGERGEEGPGAGERGEGPPGPAGPGEAGKPGEE 300
        |||||||||||||||:|           ::     |     ||||||||||||||:
Db  602 GPPGAVGPAGKDGEAGAQGPPGPAGERGEQGPAGSPGFQLPGPAGPGEAGKPGEQ 661

Qy  301 GVPGDLGAPGPSGARGEPGPPGERGVEGPPGPAGGDGAKGDAGAPGAPGSE 360
        ||||||||||||||||   ::|||||:||||||      ||||||||||:
Db  662 GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQ 721

Qy  361 GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGSE 420
        ||||| |:|||||||||||||||||||||||||||||||||||||||||||||||||:
Db  722 GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDX 781

Qy  421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGDADGEPGAKGEPGDAGKGDAGPP 480
        |||||||||||||||||||||||||||||||||||||:|| ||||||||||||||||||
Db  782 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGDAGQPGAKGEPGDAGKGDAGPP 841

Qy  481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPPGPPGPA 540
        ||||||||||||:||||||||||||||||||||||||||||||||||||:|||||||||
Db  842 GPAGPAGPPGPIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPA 901

Qy  541 GKEG 544
        ||||
Db  902 GKEG 905
```

RESULT 2
US-10-060-036-159
; Sequence 159, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
;   APPLICANT: Benson, Darin R.
;   APPLICANT: Kalos, Michael D.
;   APPLICANT: Lodes, Michael J.
;   APPLICANT: Persing, David H.
;   APPLICANT: Hepler, William T.
;   APPLICANT: Jiang, Yuqiu
;   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;   TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
;   FILE REFERENCE: 210121.566
;   CURRENT APPLICATION NUMBER: US/10/060,036
;   CURRENT FILING DATE: 2002-01-30
;   NUMBER OF SEQ ID NOS: 4560
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 159
;   LENGTH: 1464
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-060-036-159

Query Match   96.6%; Score 2966; DB 14; Length 1464;
Best Local Similarity   95.6%; Pred. No. 5.4e-146;
Matches 520; Conservative   21; Mismatches   3; Indels   0; Gaps   0;

```
Qy    1 GSEGPEGVRGEPGPPGPAGAAGPAGDGEPGAKGADGAPGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE 60
        |||||:|||||||||||||||||||  |||||||| ||||||||||||||||||||||||||||||||||||
Db  362 GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ 421

Qy   61 GPGGGPGPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEGKPGARGEPGPTGLP 120
        |||||||||||:||||||||||||||||||||||||:||||||||| |:||||||||||
Db  422 GPGGGPGPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEEGKRGARGEPGPTGLP 481

Qy  121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGKGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  482 GPPGERGGPGSRGFPGADGVAGPKGPAGERGKGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 541
```

RESULT 3
US-10-171-311-36
; Sequence 36, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
;   APPLICANT: Schlegel, Robert
;   APPLICANT: Chen, Yan
;   APPLICANT: Zhao, Xumei
;   APPLICANT: Monahan, John
;   APPLICANT: Kamatkar, Shubhangi
;   APPLICANT: Glatt, Karen
;   APPLICANT: Gannavarapu, Manjula
;   APPLICANT: Hoersh, Sebastian
;   TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;   TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
;   TITLE OF INVENTION: OF CERVICAL CANCER
;   FILE REFERENCE: MRI-035
;   CURRENT APPLICATION NUMBER: US/10/171,311
;   CURRENT FILING DATE: 2002-06-12
;   PRIOR APPLICATION NUMBER: US 60/298,159
;   PRIOR FILING DATE: 2001-06-13
;   PRIOR APPLICATION NUMBER: US 60/299,155
;   PRIOR FILING DATE: 2001-06-13
;   PRIOR APPLICATION NUMBER: US 60/335,936
;   PRIOR FILING DATE: 2001-11-14
;   NUMBER OF SEQ ID NOS: 238
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 36
;   LENGTH: 1464
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-171-311-36

Query Match   96.6%; Score 2966; DB 14; Length 1464;
Best Local Similarity   95.6%; Pred. No. 5.4e-146;
Matches 520; Conservative   21; Mismatches   3; Indels   0; Gaps   0;

```
Qy    1 GSEGPEGVRGEPGPPGPAGAAGPAGDGEPGAKGADGAPGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE 60
        |||||:|||||||||||||||||||  |||||||| ||||||||||||||||||||||||||||||||||||
Db  362 GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ 421

Qy   61 GPGGGPGPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEGKPGARGEPGPTGLP 120
        |||||||||||:||||||||||||||||||||||||:||||||||| |:||||||||||
Db  422 GPGGGPGPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEEGKRGARGEPGPTGLP 481
```

```
Qy    121 GPPCERGGPGPSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
              |||||||||||:||||||||||||||||||:||||||||||||||||:||||||||||||
Db    482 GPPGERGGPGPSRGFPGADGVAGPKGPAGERGSPGEAGRPGEAGLPGAKGLT 541

Qy    181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGPGARGEAGVMGFPGPKGAAGEPGKAGERGVP 240
              ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db    542 GSPGSPGPDGKTGPPGPAGQDGRPGPPGPGARGQAGVMGFPGPKGAAGEPGKAGERGVP 601

Qy    241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEEGPPGPGEGLPGPAGPPGEAGKPGEE 300
              ||||||:|||||||:||||||||||||||:||||||||||||||:|||||||||:
Db    602 GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQ 661

Qy    301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPGADGAPGDDGAKGDAGAPGAPGSE 360
              ||||:||||||||||||:|||||:|||||||||||||||:||||||:||||||:
Db    662 GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQ 721

Qy    361 GAPGLEGMPGERGAAGLPGPKGDRDGDAGPKGDADGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
              ||||||||||||||||:||||||||||||||||||||||||||:||||||||:
Db    722 GAPGLQGMPGERGAAGLPGPKGDRDGDAGPKGDADGSPGKDGVRGLTGPIGPPGPAGAPGDK 781

Qy    421 GESGPSGPAGPTGARGAPGDRGEPGPPGPGADGADGAPGDDGAKGDAGAPGAPGSE 480
              |||||:||||||:|||||||||||||:||||||||||||||||||:|||||:
Db    782 GESGPSGPAGPTGARGAPGDRGEPGPPGPGADGADGAPGDDGAKGDAGAPGDAGPP 841

Qy    481 GPAGPAGPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPGSGDAGPPGPPGPA 540
              ||||||||||||:||||||||||||||||||:|||:||||||:|||||||||||||:
Db    842 GPAGPAGPGPIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPGSGNAGPPGPPGPA 901

Qy    541 GKEG 544
              ||||
Db    902 GKEG 905
```

```
RESULT 4
US-10-216-705-21
; Sequence 21, Application US/10216705
; Publication No. US20030096973A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics, S.A.
; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Me
; TITLE OF INVENTION: obtaining Such and Their Uses
; FILE REFERENCE: 1149-3 DIV
; CURRENT APPLICATION NUMBER: US/10/216,705
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 09/331,347
; PRIOR FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
;   LENGTH: 1464;
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-216-705-21

Query Match         96.6%;  Score 2966;  DB 14;  Length 1464;
Best Local Similarity   95.6%;  Pred. No. 5.4e-146;
Matches   520;  Conservative   21;  Mismatches   3;  Indels   0;  Gaps   0;

Qy      1 GSEGPEGVRGEPGPPGPAGAAGPAGDGEPGAKGADGAPGIAGAPGFPGARGPSGPE 60
             ||||:||||||||||||||||||||||||||||||:||||||||:||||||||||:
Db    362 GSEGPQGVRGEPGPPGPAGAAGPAGNGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ 421

Qy     61 GPPGPPGKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEBGKPGARGEPGPTGLP 120
             ||||||||||:|||||||||||||||||||||||||||||||||||||||||||:
Db    422 GPPGPPGKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEEGKRGARGEPGPTGLP 481

Qy    121 GPPGERGGPGPSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
             ||||:||||||||||||||||||||||||||||:|||||||||||:||||||||||||
Db    482 GPPGERGGPGPSRGFPGADGVAGPKGPAGERGSPGEAGRPGEAGLPGAKGLT 541

Qy    181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGPGARGEAGVMGFPGPKGAAGEPGKAGERGVP 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGPGARGEAGVMGFPGPKGAAGEPGKAGERGVP 240
```

```
RESULT 5
US-10-149-352-2
; Sequence 2, Application US/10149352
; Publication No. US20030105050A1
; GENERAL INFORMATION:
; APPLICANT: Beri, Rajinder
; TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES
; FILE REFERENCE: 06275-254US1
; CURRENT APPLICATION NUMBER: US/10/149,352
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: PCT/GB00/04741
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: GB 9929487.8
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 4.0
; SEQ ID NO 2
;   LENGTH: 1464;
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-149-352-2

Query Match         96.6%;  Score 2966;  DB 14;  Length 1464;
Best Local Similarity   95.6%;  Pred. No. 5.4e-146;
Matches   520;  Conservative   21;  Mismatches   3;  Indels   0;  Gaps   0;

Qy      1 GSEGPEGVRGEPGPPGPAGAAGPAGDGEPGAKGADGAPGIAGAPGFPGARGPSGPE 60
             ||||:||||||||||||||||||||||||||||||:||||||||:||||||||||:
Db    362 GSEGPQGVRGEPGPPGPAGAAGPAGNGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ 421

Qy     61 GPPGPPGKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEBGKPGARGEPGPTGLP 120
             ||||||||||:|||||||||||||||||||||||||||||||||||||||||||:
Db    422 GPPGPPGKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEEGKRGARGEPGPTGLP 481

Qy    121 GPPGERGGPGPSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
             ||||:||||||||||||||||||||||||||||:|||||||||||:||||||||||||
Db    482 GPPGERGGPGPSRGFPGADGVAGPKGPAGERGSPGEAGRPGEAGLPGAKGLT 541

Qy    181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGPGARGEAGVMGFPGPKGAAGEPGKAGERGVP 240
             ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db    542 GSPGSPGPDGKTGPPGPAGQDGRPGPPGPGARGQAGVMGFPGPKGAAGEPGKAGERGVP 601

Qy    241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEEGPPGPGEGLPGPAGPPGEAGKPGEE 300
             ||||||:|||||||:||||||||||||||:||||||||||||||:|||||||||:
Db    602 GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQ 661
```

```
Qy  301 GVPGDLGAPGPSGARGEPGFPGERGVEGPGPGAGPGAGDGEPGPGAGPGAGDGDGAKGDAGAPGAPGSE 360
        |||||||||||||||||||  |:|||||:|||||||||:||||||:||| :||||||||||||
Db  662 GVPGDLGAPGPSGARGERGFPGERGVQGPGPGAGPGPGANGAPGNDGAKGDAGAPGAPGSQ 721

Qy  361 GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
        ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  722 GAPGLQQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 781

Qy  421 GESGPSGPAGVTGARGAPGDRGEPGPPGPAGFAGPPGADGEPGAKGEPGDAGAKGDAGPP 480
        |||||||||| |||||||||||||||||||| ||||||||:|||||||:|||||||||||
Db  782 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPPGADGQPGAKGEPGDAGAKGDAGPP 841

Qy  481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGDATGFPGAAGRVGPPGPGSDAGPPGPPGPA 540
        |||||||||||||:|||||||||||||||||||||||||||||||||: |:||||||||||
Db  842 GPAGPAGPPGPIGNVGAPGAKGARGSAGPPGDATGFPGAAGRVGPPGPSNAGPPGPPGPA 901

Qy  541 GKEG 544
        ||||
Db  902 GKEG 905
```

RESULT 6
US-10-177-293-65
; Sequence 65, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 65
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-65

Query Match   96.6%; Score 2966; DB 14; Length 1464;
Best Local Similarity   95.6%; Pred. No. 5.4e-146;
Matches 520; Conservative 21; Mismatches 21; Indels 3; Gaps 0;
```

```
Qy    1 GSEGPEGVRGEPGPPGPAGAAGAGDPGADGEPGAKGADGAPGIAGADGFPGARGPSGPE 60
        |||||:||||||||||||||||| |||:||||||||||||:||||||:|||||||||:
Db  362 GSEGPQGVRGEPGPPGPAGAAGPAGNGPGADGAPGAGNGAPGIAGAPGFPGARGPSGPQ 421

Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEGKPGARGEPGPTGLP 120
        ||||||||||:||||||||||||||||||||||||:||||||||||:||||||||||||
Db  422 GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEGKRGARGEPGPTGLP 481

Qy  121 GPPGERGGPSRGFPGADGVAGPKGPAGERGSPPGAGERGPGEAGLPGAKGLT 180
        |||||||||||||||||||||||||||||:||| |  ||||||||||||||
Db  482 GPPGERGGPSRGFPGADGVAGPKGPAGERGSPGAGRPGEAGLPGAKGLT 541

Qy  181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGPAGRGEAGVWGFPGPKGAAGEPGKAGERGVP 240
        ||||||||||||||||||||:|||||||||||||||||||  :||||||||||||||||||||
Db  542 GSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGPAGRGEAGQAGVWGFPGPKGAAGEPGKAGERGVP 601

Qy  241 GPPGAVGPAGKDGEAGAGPPGPGPAGPAGEGKRGEEGPAGSPGFEGLPGPAGPPGEAGKPGEE 300
        |||||||||||||||||||| ||:||||||||||||||||||| ||||:||||||||||||||
Db  602 GPPGAVGPAGKDGEAGAQGPPGPAGPAGEGKQGEEGPAGSPGFQGLPGPAGPPGEAGKPGEQ 661

Qy  301 GVPGDLGAPGPSGARGEPGFPGERGVEGPGPGAGPGAGDGEPGPGAGPGAGDGDGAKGDAGAPGAPGSE 360
        |||||||||||||||||  |:|||||:|||||||||:||||||:||| :||||||||||||
Db  662 GVPGDLGAPGPSGARGERGFPGERGVQGPGPGAGPGPGANGAPGNDGAKGDAGAPGAPGSQ 721

Qy  361 GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
        ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  722 GAPGLQQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 781

Qy  421 GESGPSGPAGVTGARGAPGDRGEPGPPGPAGFAGPPGADGEPGAKGEPGDAGAKGDAGPP 480
        |||||||||| |||||||||||||||||||| ||||||||:|||||||:|||||||||||
Db  782 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPPGADGQPGAKGEPGDAGAKGDAGPP 841

Qy  481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGDATGFPGAAGRVGPPGPGSDAGPPGPPGPA 540
        |||||||||||||:|||||||||||||||||||||||||||||||||: |:||||||||||
Db  842 GPAGPAGPPGPIGNVGAPGAKGARGSAGPPGDATGFPGAAGRVGPPGPSNAGPPGPPGPA 901

Qy  541 GKEG 544
        ||||
Db  902 GKEG 905
```

RESULT 7
US-10-301-822-28
; Sequence 28, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MPM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/10/301,822
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-301-822-28

```
Query Match              96.6%;   Score 2966;  DB 15;   Length 1464;
Best Local Similarity    95.6%;   Pred. No. 5.4e-146;
Matches 520;   Conservative 21;   Mismatches 3;   Indels 0;   Gaps 0;

Qy     1  GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGFPPGARGPSGPE   60
          ||||:|||||||||||||||||||||:|||||||:|||||||||||||||| :
Db   362  GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ   421

Qy    61  GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGEGPPGPAGEGKPGARGEPGPTGLP    120
          ||||||||:||||||||||||||||||||||||:  ||||||||:||||||||||||
Db   422  GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGQGPPGPAGEGKRGARGEPGPTGLP    481

Qy   121  GPPGERGGPGSRGFPGADGVAGPKGPAGKGSPGEAGRPGEAGLPGAKGLT           180
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   482  GPPGERGGPGSRGFPGADGVAGPKGPAGKGSPGEAGRPGEAGLPGAKGLT           541

Qy   181  GSPGSPGDGKTGPPGPAGEDGRPGPPGPPGPAGEDGRPGPPGPPGPAGEAGVMGFPPGPKGAAGEPGKAGERGVP   240
          |||||||||||||||||:||||||||||||||||||| : ||||||||:||||||||||||||||||
Db   542  GSPGSPGDGKTGPPGPAGQDGRPGPPGPPGPAGRQAGVMGFPGPKGAAGEPGKAGERGVP           601

Qy   241  GPPGAVGPAGKDGEAGAEGPPGPAGERGEEGPAGSPGFEGLPGPAGPPGPEAGKPGEE   300
          ||||||||||||||||:|||||||||||:|||||||||:||||||||||||||||:
Db   602  GPPGAVGPAGKDGEAGAQGPPGPAGERGEQGPAGSPGFQGLPGPAGPPGPEAGKPGEQ   661

Qy   301  GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGPADGAPGDDGAKGDAGAPGAPGSE   360
          ||||||||||||||||:|||||||||:||||||||   :||||||||||||||||||:
Db   662  GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQ   721

Qy   361  GAPGLEGMPGERGAAGLPGPKGDRGDAGSPGKDGVRGLTGPIGPPGPAGAPGDK       420
          |||||:||||||||||||||||||||||  :|||||||||||||||||||||||
Db   722  GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADSPGKDGVRGLTGPIGPPGPAGAPGDK   781

Qy   421  GESGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGPAGDGEPGAKGEPGDAGAKGDAGPP   480
          |||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db   782  GESGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGPAGDGQPGAKGEPGDAGAKGDAGPP   841

Qy   481  GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGPGPPGPPGPA                     540
          ||||||||||||:||||||||||||||||||:         ||||||||
Db   842  GPAGPAGPPGPIGNVGAPGAKGARGSAGPPGDATGFPGAAGRVGPPGPSGNAGPPGPPGPA   901

Qy   541  GKEG   544
          ||||
Db   902  GKEG   905
```

RESULT 8
US-10-291-265-243
; Sequence 243, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: NO. US20030232054A1e1 Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 243
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-243

RESULT 9
US-10-402-089-2
; Sequence 2, Application US/10402089
; Publication No. US20040005563A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.3 CON
; CURRENT APPLICATION NUMBER: US/10/402,089
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 1463
; TYPE: PRT
; ORGANISM: Bos Taurus
US-10-402-089-2

```
Query Match              95.6%;   Score 2936;  DB 15;   Length 1463;
Best Local Similarity    94.1%;   Pred. No. 1.9e-144;
Matches 512;   Conservative 25;   Mismatches 7;   Indels 0;   Gaps 0;

Qy     1  GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE   60
```

Db 421 GPGSPPGPKGNSGEPGPAPGSKDTGAKGEPGPTGIQGPPGPAGEEGKRGARGEPGPAGLP 480
        |||| :: ||||||||||:||||:||||||:||||||||:|||||||| : :|||||
Qy 121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
        |||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db 481 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 540
Qy 181 GSPGSPGPGKTGPPGPAGEDGRPGPPGPAGEAGVMGFPGPKGAAGEPGKAGERGVP 240
        ||||||||||||||||:||||||||||||||:|||||||||||||||||||||||
Db 541 GSPGSPGPGKTGPPGPAGQDGRPGPPGPAGQAGVMGFPGPKGAAGEPGKAGERGVP 600
Qy 241 GPPGAVGPAGKDGEAGABGPPGPAGAGERGEEGPAGSPGFEGLPGPAGPPGEAGKPGEE 300
        ||||||||||||| |:|||||||||||:||:||||||||:||||||||||||||||:
Db 601 GPPGAVGPAGKDGAGAQGPPGPAGAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQ 660
Qy 301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPGDDGAKGDAGDAGAPGAPGSE 360
        |||||||||||||||||:|||||||:||||||||||::||||||||:|||||||:
Db 661 GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQ 720
Qy 361 GAPGLEGMPGERGAAGLPGPKGDRDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
        ||||| ||||||||||||||||||:|||||||| :|||||||||||||||||||||||
Db 721 GAPGLQGMPGERGAAGLPGPKGDRDAGPKGADGAPGKDGVRGLTGPIGPPGPAGAPGDK 780
Qy 421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPPGDAGBEPGDAGAKGDAGPP 480
        ||:|||||||||||||||||||||||||||||||||| :|||||||||||||||
Db 781 GEAGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPPGDADGQPGAKGEPGDAGAKGDAGDP 840
Qy 481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGDATGFPGAAGRVGPPGPSGDAGPPPGPA 540
        ||||||||||||:|:||||||||||||||||||||||||||||||||||||:||||||
Db 841 GPAGPAGPPGPIGNVGAPGPKGARGSAGPPGDATGFPGAAGRVGPPGPSGNAGPPGPPGPA 900
Qy 541 GKEG 544
        ||||
Db 901 GKEG 904

RESULT 11
US-10-342-331-49
; Sequence 49, Application US/10342331
; Publication No. US20030222205A1
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOUWSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHELE D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/10/342,331
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: US/09/219,849
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 49
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid sequence
US-10-342-331-49

Query Match   93.1%; Score 2857; DB 15; Length 822;
Best Local Similarity 91.4%; Pred. No. 1.5e-140;
Matches 497; Conservative 28; Mismatches 19; Indels 0; Gaps 0;

Qy   1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE 60
        ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||

Db 361 GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ 420
        ||||:|||||||||||||||||||||:||||||||||||||||||||||||||||||||:
Qy   61 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 120
        |||||||||||||||||||||:|||||||||||||||||||||||||:|:|||||||||
Db 421 GPGSPPGPKGNSGEPGPAPGSKGDTGAKGEPGPTGIQGPPGPAGEEGKRGARGEPGPAGLP 480
Qy 121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
        |||||||||||||||||||||:|||||||||||||||||||||||||:|:|||||||||
Db 481 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 540
Qy 181 GSPGSPGPGKTGPPGPAGEDGRPGPPGPAGEAGVMGFPGPKGAAGEPGKAGERGVP 240
        ||||||||||||||||:||||||||||||||:|||||||||||||||||||||||
Db 541 GSPGSPGPGKTGPPGPAGQDGRPGPPGPAGQAGVMGFPGPKGAAGEPGKAGERGVP 600
Qy 241 GPPGAVGPAGKDGEAGABGPPGPAGAGERGEEGPAGSPGFEGLPGPAGPPGEAGKPGEE 300
        ||||||||||||| |:|||||||||||:||:||||||||:||||||||||||||||:
Db 601 GPPGAVGPAGKDGAGAQGPPGPAGAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQ 660
Qy 301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPGDDGAKGDAGDAGAPGAPGSE 360
        |||||||||||||||||:|||||||:||||||||||::||||||||:|||||||:
Db 661 GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQ 720
Qy 361 GAPGLEGMPGERGAAGLPGPKGDRDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
        ||||| ||||||||||||||||||:|||||||| :|||||||||||||||||||||||
Db 721 GAPGLQGMPGERGAAGLPGPKGDRDAGPKGADGAPGKDGVRGLTGPIGPPGPAGAPGDK 780
Qy 421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPPGDAGBEPGDAGAKGDAGPP 480
        ||:|||||||||||||||||||||||||||||||||| :|||||||||||||||
Db 781 GEAGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGDADGQPGAKGEPGDAGAKGDAGPP 840
Qy 481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGDATGFPGAAGRVGPPGPSGDAGPPPGPA 540
        ||||||||||||:|:||||||||||||||||||||||||||||||||||||:||||||
Db 841 GEAGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGDADGQPGAKGEPGDAGAKGDAGPP 900
Qy 541 GKEG 544
        ||||
Db 901 GKEG 904

RESULT 10
US-10-402-072A-2
; Sequence 2, Application US/10402072A
; Publication No. US20040018592A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.2 CON
; CURRENT APPLICATION NUMBER: US/10/402,072A
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 1463
; TYPE: PRT
; ORGANISM: Bos Taurus
US-10-402-072A-2

Query Match   95.6%; Score 2936; DB 15; Length 1463;
Best Local Similarity 94.1%; Pred. No. 1.9e-144;
Matches 512; Conservative 25; Mismatches 7; Indels 0; Gaps 0;

Qy   1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE 60
        ||||:|||||||||||||||||||||:||||||||||||||||||||||||||||||||:
Db 361 GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ 420
Qy   61 GPGGPPPGKGDSGEPGPAPGSKGDTGAKGEPGPAGEGKPGARGEPGPTGLP 120
        || |||||||||||||||||||||||||||||| :||||:||||||||

Db 183 GSEPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPGSPQ 242
           |||| ||||||||||||||||||||| ||||| :||||||||||||||||||||||::|
Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEEGKPGARGEPGPTGLP 120
           |||||||||:|:||||||||||||||||||||| ||||||||||||:||||||||||
Db 243 GPSGPPGPKGNSGEPGAPGNKGDTGAKGEPGATGVQGPPGPAGEEGKRGARGEPGPSGLP 302

Qy 121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGERGSPGPAGERGPGEAGLPGAKGLT 180
           ||||||||||||||||||||| |||:||||||||||||||||:|||||||||||||||
Db 303 GPPGERGGPGSRGFPGADGCVAGPKGPSGBERGAPGPAGERGQGPGEAGLPGAKGLT 362

Qy 181 GPSGSPGDGKTGPPGPAGEDGRPGPPGPPGARGEAGVMGFPGPKGAAGEPGKAGERGVP 240
           |:||:|||||||||||||||:|||||||||||||||||:||||||||||||||||||
Db 363 GSSGSPGDGKTGPPGPAGQDGRPGPAGPAGPKGPSGBERGAPGPAGERGQGPGEAGLP 422

Qy 241 GPPGAVGPAGKDGEAGABGPPGPAGPAGEGPAGSPGFEGLPGAGPPGPGEAGKPGEE 300
           ||||||||||||||||||::||||||||||||||:|||||||||||||||||||||:
Db 423 GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGFQGLPGAGPPGEAGKPGEQ 482

Qy 301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPGADAPGDDGAKGDAGAPGSE 360
           :||||||||||||||||||::|||||||||| ::|||:|||:|||:|||||:
Db 483 GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGNNGAPGNDGAKGDTGAPGSQ 542

Qy 361 GAPGLEGMPGERGAAGLPGPKGDRDGDAGPKGDGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
           |||||:|||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db 543 GAPGLQGMPGERGAAGLPGPKGDRDGDAGPKGDGSPGKDGARGLTGPIGPPGPAGAPGDK 602

Qy 421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPPGPAGPAGPPGPAGPAGPP 480
           |:|||:||||||||||:|||:|||||||||||||||||||||||||||||::||
Db 603 GEAGPSGPGPTGARGAPGDRGEAGPPGPAGPAGPPGPAGPAGPPGPGPSGNAGPPGPPGPPGPV 662

Qy 481 GPAGPAGPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPPGPPGPA 540
           |:|||||||||| ||||||||||||||||||||||||||||||||||||||||::|||
Db 663 GPAGPAGPGPIGNVGAPGPGNVGADGPKGPRGAAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPV 722

Qy 541 GKEG 544
           ||||
Db 723 GKEG 726

RESULT 12
US-10-402-089-8
; Sequence 8, Application US/10402089
; Publication No. US20040005663A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.3 CON
; CURRENT APPLICATION NUMBER: US/10/402,089
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
;   LENGTH: 1449
;   TYPE: PRT
;   ORGANISM: Sus scrofa
US-10-402-089-8

Query Match   91.7%; Score 2816.5; DB 15; Length 1449;
Best Local Similarity   91.0%; Pred. No. 2.9e-138;
Matches   495; Conservative 22; Mismatches 10; Indels 17; Gaps 1;

Qy   1 GSEGPEGVRGEPGPPGPAGAAGPAGDGPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE 60
           |:||::|||||||||||||||||| |||:||||:|||||:||||||||||||||||||
Db 364 GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGGKGANGAPGIAGAPGFPGARGPSGPQ 423

Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEEGKPGARGEPGPTGLP 120
           |||||||||:|:|||||||||||||||||||| ||||||||||||:||||||||||
Db 424 GPSGPPGPKGNSGEPGAPGSKGDTGAKGEPGPTGVQGPPGPAGEEGKRGARGEPGPAGLP 483

Qy 121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGERGSPGPAGERGPGEAGLPGAKGLT 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 484 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGERGSPGPAGERGPGEAGLPGAKGLT 543

RESULT 13
US-10-402-072A-8
; Sequence 8, Application US/10402072A
; Publication No. US20040018592A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.2 CON
; CURRENT APPLICATION NUMBER: US/10/402,072A
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
;   LENGTH: 1449
;   TYPE: PRT
;   ORGANISM: Sus scrofa
US-10-402-072A-8

Query Match   91.7%; Score 2816.5; DB 15; Length 1449;
Best Local Similarity   91.0%; Pred. No. 2.9e-138;
Matches   495; Conservative 22; Mismatches 10; Indels 17; Gaps 1;

Qy 121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGERGSPGPAGERGPGEAGLPGAKGLT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 484 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGERGSPGPAGERGPGEAGLPGAKGLT 543

Qy 181 GSPGSPGDGKTGPPGPPGAGEDGRPGEPGPPGPPGARGEAGVMGFPGPKGAAGEPGKAGERGVP 240
           |:||:|||||||||||||||||||||||||||||||:||||:|||||||||||||||||||
Db 544 GSPGSPGDGKTGPPGPAGQDGRPGPAGRGQAGVMGFPGPKGAAGEPGKAGERGVP 603

Qy 241 GPPGAVGPAGKDGEAGABGPPGPAGPAGEGREGPAGSPGFEGLPGAGPPGPGEAGKPGEE 300
           ||||||||||||||||: ||||||||||||||::||||||:|||||||||||||||||:
Db 604 GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPGLPGAGPPGQGLPGAGPPGEAGKPGEQ 663

Qy 301 GVPGDLGAPGPSGARGEPGPFPGERGVEGPPGRGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGSE 360
           :||||||||||||||||||:::||||||||||||||||||| :|||:|||::|||:||:|||||:
Db 664 GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGSQ 723

Qy 361 GAPGLEGMPGERGAAGLPGPKGDRDGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
           |||||:|||||||||||||||||:||||||:||||||||||||||||||||||||||||||
Db 724 GAPGLQGMPGERGAAGLPGPKGDRDGDAGPKGADGPKGDGVRGLTGPIGPPGPAGAPGDK 783

Qy 421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPPGPAGPAGPPGDAGKGDGAPP 480
           |:|||:||||||||||||||||:||||||||||||||||||||||||------------
Db 784 GETGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPPGPAGDGQPGAKG-------------- 829

Qy 481 GPAGPAGPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPPGPPGPA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 830 --GPTGPPGPIGSVGAPGPKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPA 886

Qy 541 GKEG 544
           ||||
Db 887 GKEG 890

```
Qy  181 GSPGSPGPDGKTGPPGPGAGBDGRPGPPGPPGARGBAGVMGFPPGPKGAAGEPGKAGERGVP 240
            ||||||||||||||||||| |||||||||||   ||||||||||||||||||||
Db  544 GSPGSPGPDGKTGPPGPGAGQGDGRPGAGPGPPGPPGCARGQAGVMGFPPGPKGAAGEPGKAGERGVP 603

Qy  241 GPPGAVGPAGDGEAGAEGPPGPAGBRGEEGPAGSPGPEGLPGPAGPPGEAGKPGEE 300
            |||||||| |||||| ||||||||: :|||:   |||||| ||||||||||||
Db  604 GPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQ 663

Qy  301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPGPAGPGDDGAKGDAGAPGAPGSE 360
            ||||||||||||||| |||||||||:|||||||||| |||:|||||||||||||||:
Db  664 GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPGANGAPGNDGAKGDAGAPGAPGSQ 723

Qy  361 GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGCVRGLTGPIGPPGPAGAPGDK 420
            |||| |||:||||||||||||:|||:|||||||||||| |||||||||||||||||||||
Db  724 GAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGAPGKDGVRGLTGPIGPPGPAGAPGDK 783

Qy  421 GESGPSGPAGPTGEARGAPGDRGEPGPPGPAGPAGPPGDAGEPGAKGEPGDAGAKDGAGPP 480
            ||:||:|||||||||||||||:||||||||||||||:|:|||| |:|||:|||
Db  784 GETGPSGPAGPTGEARGAPGDRGEPGPPGPAGPFAGPPGDADGQPGAKG------------- 829

Qy  481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAACRVGPPGPGPSGDAGPPGPPGPA 540
            ||||||||||||||||||||:|||||||||||:||||||||:||||||||||||:|||||||
Db  830 ---GPTGPPGPIGSVGAPGPKGARGSAGPPGATGFPGAACRVGPPGPGPSGNAGPPGPPGPA 886

Qy  541 GKEG 544
            ||||
Db  887 GKEG 890
```

RESULT 14
US-10-058-124-18
; Sequence 18, Application US/10058124
; Publication No. US20030119058A1
; GENERAL INFORMATION:
;   APPLICANT: Qvist, Per
;   APPLICANT: Bonde, Martin
;   TITLE OF INVENTION: A Method for Assaying Collagen Fragments
;   in Body Fluids, A Test Kit and Means for Carrying Out the
;   Method and Use of the Method to Diagnose the Presence of
;   Disorders Associated with the Metabolism of
;   NUMBER OF SEQUENCES: 21
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Darby & Darby PC
;   STREET: 805 Third Avenue
;   CITY: New York
;   STATE: New York
;   COUNTRY: USA
;   ZIP: 10022
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/058,124
;   FILING DATE: 29-Jan-2002
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 09/570,573
;   FILING DATE: 2002-MAY-12
;   APPLICATION NUMBER: 08/187,319
;   FILING DATE: <Unknown>
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Gogoris, Adda C
;   REGISTRATION NUMBER: 29,714
;   REFERENCE/DOCKET NUMBER: 4305/08701
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 212-527-7700
;   TELEFAX: 212-753-6237
;   TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1341 amino acids

;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   ORIGINAL SOURCE:
;   ORGANISM: Homo sapiens
;   IMMEDIATE SOURCE:
;   CLONE: COLLAGEN ALPHA 1 (I)
;   SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-10-058-124-18

Query Match   90.9%; Score 2791.5; DB 14; Length 1341;
Best Local Similarity   87.6%; Pred. No. 5.3e-137;
Matches 500; Conservative 21; Mismatches 23; Indels 27; Gaps 4;

```
Qy    1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGE---PGAKGADGAPGIAGAPGFPGAR--- 54
           ||||||| |||||||||||||||||||||||   ||:|| :: ||  ||| | |||
Db  211 GASGVMGPFGPFGPGPGKBGBBGZAGKPGRPGZRGPGPZGARGLPGTAGLPGMKGHRGFS 270

Qy   55 -------GPSGPEG-------PG--GPPGPKGDSGEPGAPGSKGDTGAKGEPGPV 93
                  ||:|||       :  :|||||||||||||:|||||||||:|||||||
Db  271 GLBGAKGBAGPAGPKGZPGSPGZBGAPGZMGFPGPGKGNSGEPGAPGSKGDTGAKGEPGPV 330

Qy   94 GVEGPPGPAGEEGKPGARGEPGPTGLPGPPGERGGPGSRGFPGADGVAGPKGPAGERGSP 153
           :|:||||||||| ||||||||||||||||||||::|||||||||||||||||||||||
Db  331 GVQGPPGPAGEEGKRGARGEPGPTGLPGPPGERGGPGSRGFPGADGVAGPKGPAGERGSP 390

Qy  154 GPAGPKGSPGEAGRPGEAGLPGAKGLTGSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGAR 213
           |||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db  391 GPAGPKGSPGEAGRPGEAGLPGAKGLTGSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGAR 450

Qy  214 GEAGVMGFPGPKGAAGEPGKAGERGVPGPPGAVGPAGKDGEAGBGPPGPAGEAGBGPPGPAGERGEE 273
           | |||:|||||||||||||:|||||||||||||| ||||||:|||||||||||||||||||||
Db  451 GQAGVMGFPGPKGAAGEPGKAGERGVPGPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQ 510

Qy  274 GPAGSPGFEGLPGPAGPPGEAGKPGEEGVPGDLGAPGPSGARGEPGFPGERGVEGPPGPA 333
           |||||||:|||||||||||||||||||:|||||||||||||||||||||||||:||||||
Db  511 GPAGSPGFQCLPGPAGPPGEAGKPGEQGVPGDLGAPGPSGARGERGFPGERGVQGPPGPA 570

Qy  334 GPPGDGAPGDDGAKGDAGAPGAPGSEGAPGLEGMPGERGAAGLPGPKGDRGDAGPKGAD 393
           |||| ||||:|||||||||||||||:| ||| ||:||||||||||||:|||:|||||||
Db  571 GPPGANGAPGNDGAKGDAGAPGAPGSQGAPGLQGMPGERGAAGLPGPKGDRGDAGPKGAD 630

Qy  394 GSPGKDGVRGLTGPIGPPGPAGAPGDKGESGPSGPAGPTGEARGAPGDRGEPGPPGPAGFA 453
           |||||||||||||||||||||||||||||:||:|||||||||||||||:|||||||||||
Db  631 GSPGKDGVRGLTGPIGPPGPAGADGKGSGPSGPAGPTGARGARGDRGEPGPPGPAGFA 690

Qy  454 GPPGADGEPGAKGEPGDAGAKDGAGPPGPGPAGPAGPPGPIGDVGAPGAKGARGSAGPPGAT 513
           ||||||:|:|||:|||||||:|||| ||||:||||:|:|||||||||||:||||||||||||
Db  691 GPPGADGQPGAKGEPGDAGAKDGAGPPGPGPAGPGPAGPGPIGNVGAPGPIGNVGAPGAT 750

Qy  514 GFPGAAGRVGPPGPSGDAGPPGPPGPAGKEG 544
           |||||||||||||| ||:||||||||||||
Db  751 GFPGAAGRVGPPGPSGNAGPPGPPGPAGKEG 781
```

RESULT 15
US-10-194-441A-1
; Sequence 1, Application US/10194441A
; Publication No. US20030148944A1
; GENERAL INFORMATION:
;   APPLICANT: Holmdahl, Rikard
;   APPLICANT: Engstrom, Jan Ake
;   APPLICANT: Kihlberg, Jan
;   APPLICANT: Burkhardt, Harald
;   TITLE OF INVENTION: TRIPLE POLYPEPTIDE COMPLEXES
;   FILE REFERENCE: 1145-010001
;   CURRENT APPLICATION NUMBER: US/10/194,441A
;   CURRENT FILING DATE: 2002-07-11
;   PRIOR APPLICATION NUMBER: US 60/305,048
;   PRIOR FILING DATE: 2001-07-12
;   NUMBER OF SEQ ID NOS: 87
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1

```
;   LENGTH: 1014
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-194-441A-1

Query Match            72.0%;    Score 2209;  DB 14;   Length 1014;
Best Local Similarity  72.2%;    Pred. No. 5.8e-107;
Matches  393;  Conservative  38;  Mismatches  113;  Indels   0;  Gaps   0;

Qy     1  GSEGPEGVRGEPGPPGPGPAAAGPAGDPGADGEPGAKGADGAPGIAAGAPGFPGARGPSGPE  60
          ||||||| ||| | | |||| |  :|:|  || || ||  |||||||||| |||||| || ||:
Db   184  GPEGAQGPRGEPGTPGSPGPAGASGNPGTDGIPGAKGSAGAPGIAGAPGFPGPRGPPGPGPQ  243

Qy    61  GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPGPGPAGEGKPGARGEPGPTGLP  120
          |||||| ||| ||  | |  ||| |:  ||  ||||:| ||||||||| ||||||||||||
Db   244  GATGPLGPKGQTGEPGIAGFKGEQGPKGEPGPAGPQGAPGPAGBEGKRGARGEPGGVGPI  303

Qy   121  GPPGERGPGPGSRGPPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT  180
          ||||||| |||||  |:|||| |::|||| ||||:| ||  |: :|||||| |||||||||
Db   304  GPPGERGRAPGNRGFPGQDGLAGPKGAPGERGPSGLAGPKGANGDPGRPGEPGLPGARGLT  363

Qy   181  GSPGSPGPDGKTGPPGPAGEDGRRPGPPGPPGCARGEAGVWGFPGPKGKGAAGEPGKAGERGVP  240
          | || | || | ||| || |  ||| | || | |:| ||  |||||  | |||||| |:|:|
Db   364  GRPGDAGPQGKVGPSGAPGEDGRPGPPGQGARGQPGVWGFPGPKGANGEPGKAGEKGLP  423

Qy   241  GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEEGPAGSPGFEGLLPGPAGPPGEAGKPGEB  300
          | || | |  | ||  | |||| || |||  | :| ||| |   |:||| | || |:| |||
Db   424  GAPGLIRGLPGKDGETGAAGPPGPAGPAGERGEQGAAPGPSGFQGLLPGPGPPGBGGKQGDQ  483

Qy   301  GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPGPAGDGAKGDDGAKDAGAPGSE  360
          || ||:| ||| |  | |||  |:|| | :|:|  || || | | |||| :| |::
Db   484  GVPGBAGAAPGLIVGPRGERGFPGERGSPGAQLQGPRGLPGTPGTDGPKGASGPAGPPGAQ  543

Qy   361  GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  420
          ||||| || ||| || |||||| |  | | :|:||| | ||||||:| ||||||||  |:
Db   544  GPPGLQGMPGERGAAGIAGPKGDRGDVGBKGPEGAPGKDGGRGLTGPIGPPGPAGANGEK  603

Qy   421  GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPPGADGEPGAKGEPGDAGAKGDAGPP  480
          | |||| | |||  || |  |:|:  | ||||||||| :| |||| | | ||  ||  ||
Db   604  GEVGPPGPAGSAGARGAPGEPGETGPPGPAGFAGPPGADGQPGAKGEQGEAGQKGDAGAP  663

Qy   481  GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPPGPPGPA  540
          |:| | | | ||  | || ||  | || || ||| ||||||||| ||| |||| || ||
Db   664  GPQGPSGAPGPQGPTGVTGPKGARGAQGPPGATGFPGAAGRVGPPGSNGNPGPPGPGPS  723

Qy   541  GKEG  544
          |||:|
Db   724  GKDG  727
```

Search completed: May 7, 2004, 16:31:58
Job time : 50 secs

```
              Copyright  GenCore version 5.1.6
                      (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     May  7, 2004, 16:25:44 ; Search time 20 seconds
                                       (without alignments)
                                       2616.410 Million cell updates/sec

Title:         US-10-658-989A-4
Perfect score: 3070
Sequence:    1 GSEGGPEGVRGEPGPGPAGA........PGPSGDAGPPGPGPAGKEG 544

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:    283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:    283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   PIR_78:*
             1:  pir1:*
             2:  pir2:*
             3:  pir3:*
             4:  pir4:*

Pred. No. is the number of results predicted by chance to have a
          score greater than or equal to the score of the result being printed,
          and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	2966	96.6	1464	1	CGHU1S	collagen alpha 1(I
2	2857	93.1	1453	2	S21626	collagen alpha 1(I
3	2775	90.4	1042	1	CGCHIS	collagen alpha 1(I
4	2202	72.3	1418	2	T45467	collagen alpha 1(I
5	2202	71.7	1487	1	CGHU6C	collagen alpha 1(I
6	2189	71.3	1419	2	A41182	collagen alpha 1(I
7	2189	71.3	1487	2	B41182	collagen alpha 1(I
8	2154	70.2	671	1	CGRTIS	collagen alpha 1(I
9	2147	69.9	1486	1	B40333	collagen alpha 1(I
10	2131	69.4	1492	2	A40333	collagen alpha 1(I
11	2076.5	67.6	779	1	CGBO1S	collagen alpha 1(I
12	1921.5	62.6	1496	1	CGHU2V	collagen alpha 2(I
13	1915	62.4	1373	1	A43291	collagen alpha 2(I
14	1910	62.2	1497	2	I49607	procollagen type V
15	1908	62.1	1464	2	S59856	collagen alpha 1(I
16	1905.5	62.1	1049	1	CGBO7S	collagen alpha 1(I
17	1901	61.9	1466	1	CGHU7L	collagen alpha 1(I
18	1886	61.4	1366	1	CGHU2S	collagen alpha 2(I
19	1881	61.3	886	2	I50694	collagen alpha 1(I
20	1550.5	50.5	1414	1	S23809	collagen alpha 2(I
21	1457.5	47.5	1838	1	CGHU1V	collagen alpha 1(V
22	1444	47.0	1806	1	CGHU1E	collagen alpha 1(X
23	1432.5	46.7	1843	2	S18803	collagen alpha 1(V
24	1418.5	46.0	1027	2	S28774	collagen alpha cha
25	1413	46.0	964	1	CGCH2S	collagen alpha 2(I
26	1411.5	46.0	1024	2	S18251	collagen alpha 1(X
27	1400	45.6	888	2	S28791	collagen alpha 1(X
28	1399.5	45.6	1546	1	CGHU2E	collagen alpha 2(X
29	1376	44.8	1691	1	S22917	collagen alpha 5(I
30	1375.5	44.8	2944	2	A54849	collagen alpha 1(V
31	1372	44.7	1763	2	S16366	collagen alpha 2(I
32	1364.5	44.4	1549	2	I48103	type VII collagen
33	1357.5	44.2	1690	1	CGHU1B	collagen alpha 4(I
34	1339.5	43.6	1603	2	S23810	collagen alpha 1(X
35	1339.5	43.6	1670	1	CGHU3B	collagen alpha 3(I
36	1335.5	43.5	920	1	A45748	collagen alpha 1(V
37	1320	43.0	1669	1	CGHU4B	collagen alpha 1(I
38	1310.5	42.7	1758	2	T29350	hypothetical prote
39	1310	42.7	1669	2	CGMS4B	collagen alpha 2(I
40	1309	42.6	1759	2	T29351	collagen alpha 2(I
41	1303.5	42.5	812	1	S31521	collagen COLF1 - f
42	1286.5	41.9	1712	2	CGHU2B	collagen alpha 2(I
43	1279.5	41.7	754	2	A55267	collagen alpha 5(I
44	1277	41.6	921	2	S42617	collagen alpha 1(I
45	1276.5	41.6	1142	2	JX0369	collagen alpha 1(X

ALIGNMENTS

```
RESULT 1
CGHU1S
collagen alpha 1(I) chain precursor - human
N;Alternate names: procollagen alpha 1(I) chain
C;Species: Homo sapiens (man)
C;Date: 12-Aug-1981 #sequence_revision 04-Oct-1996 #text_change 31-Dec-2000
C;Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S1
5269; A29439; I53466; A02852; I37247
R;D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorious, P.J
Gene 67, 105-115, 1988
A;Title: Complete nucleotide sequence of the region encompassing the first twenty-five
A;Reference number: I60114; MUID:88329734; PMID:2843432
A;Accession: I60114
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-369,'L',371-589 <DAL>
A;Cross-references: GB:M20789; NID:g179593; NID:g179591; PIDN:AAB59373.1; PID:g179594
R;Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Proc
Biochem. J. 253, 919-922, 1988
A;Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human
A;Reference number: S01143; MUID:89025644; PMID:3178743
A;Accession: S01143
A;Molecule type: mRNA
A;Residues: 1-472 <TRO>
A;Cross-references: EMBL:X07884; NID:g30015; PIDN:CAA30731.1; PID:g30016; GB:M36546; NI
A;Note: submitted to the EMBL/GenBank/DDBJ databases by Prockop, D.J., 13-JUN-1988
R;Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.
Nature 310, 337-340, 1984
A;Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation o
A;Reference number: A93335; MUID:84270697; PMID:6462220
A;Accession: A93335
A;Molecule type: DNA
A;Residues: 1-58,'Q',60-181 <CHU>
A;Cross-references: EMBL:X00820; NID:g35557; PIDN:CAA25394.1; PID:g35658
R;Rossouw, C.M.S.; Vergeer, W.P.; du Plooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W
J. Biol. Chem. 262, 15151-15157, 1987
A;Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene en
A;Reference number: I55254; MUID:88033098; PMID:2822714
A;Accession: I55254
A;Status: translation not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-45 <ROS>
A;Cross-references: GB:J02829; NID:g180387; PIDN:AAA51993.1; PID:g180388
R;Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gelinas, R.E.
Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987
A;Title: Regulatory elements in the first intron contribute to transcriptional control
A;Reference number: A39943; MUID:88097389; PMID:3480516
A;Accession: A39943
A;Molecule type: DNA
A;Residues: 1-34 <BOR>
A;Cross-references: GB:J03559; NID:g180876; PIDN:AAA52052.1; PID:g53238
R;Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.
```

J. Biol. Chem. 260, 2315-2320, 1985
A;Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter a
A;Reference number: I55237; MUID:85130970; PMID:2857713
A;Accession: I55237
A;Status: translation not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-34 <CH2>
A;Cross-references: GB:M10627; NID:g180383; PIDN:AAA51992.1; PID:g553226
R;Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollist
J. Biol. Chem. 265, 6312-6317, 1990
A;Title: In vivo and in vitro noncovalent association of excised alpha1(II) amino-termina
rome, type VII.
A;Reference number: A35233; MUID:90202908; PMID:2318855
A;Accession: A35233
A;Molecule type: protein
A;Residues: 33-52 <WIR>
R;Click, E.M.; Bornstein, P.
Biochemistry 9, 4699-4706, 1970
A;Title: Isolation and characterization of the cyanogen bromide peptides from the alpha1
A;Reference number: A90567; MUID:71038825; PMID:5529814
A;Contents: CNBr0-1, CNBr2, CNBr4, CNBr5
A;Accession: B90567
A;Molecule type: protein
A;Residues: 162-198,'Z',200-201,'Z',203-206,'Z',208-209,'Z',211-228,'B',230,'BB',233,'Z'
A;Experimental source: skin
A;Note: evidence for 170-allysine
R;Baetge, B.; Notbohm, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, E
Eur. J. Biochem. 192, 153-159, 1990
A;Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle
A;Reference number: S11342; MUID:90382436; PMID:2169412
A;Accession: S11372
A;Molecule type: protein
A;Residues: 175-187;274-287,'P',289 <BAE>
A;Note: sequence of collagen alpha 1(S)(I) isolated from bone after pepsin digestion
R;Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonzalez
J. Biol. Chem. 266, 21827-21832, 1991
A;Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain
A;Reference number: I55342; MUID:92042092; PMID:1718984
A;Accession: I55342
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 258-268;1347-1357 <DEA>
A;Cross-references: GB:S67495; NID:g239007; PIDN:AAB20350.1; PID:g239008
A;Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg report
R;Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
J. Biol. Chem. 245, 5042-5048, 1970
A;Title: Comparative study of glycopeptides derived from selected vertebrate collagens.
A;Reference number: A92069; MUID:71001508; PMID:4319110
A;Accession: A92069
A;Molecule type: protein
A;Residues: 263-268 <MOR>
A;Experimental source: skin
A;Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
R;Labhard, M.E.; Hollister, D.W.
Matrix 10, 124-130, 1990
A;Title: Segmental amplification of the entire helical and telopeptide regions of the cD
A;Reference number: S15989; MUID:90326017; PMID:2374517
A;Accession: S15989
A;Molecule type: mRNA
A;Residues: 281-302;402-420;823-843;925-944;1026-1045;1143-1162 <LAB>
R;Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries, W.N
Connect. Tissue Res. 29, 1-11, 1993
A;Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of
A;Reference number: I52905; MUID:93339042; PMID:8339541

A;Accession: I52905
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 342-352,'C',354-359 <WI2>
A;Cross-references: GB:S64717; NID:g408195; PIDN:AAB27677.1; PID:g408196
A;Note: mutant sequence from patient with osteogenesis imperfecta
R;Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.
Biochemistry 22, 5213-5223, 1983
A;Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalpha1
A;Reference number: A90476; MUID:84080385; PMID:6689127
A;Accession: A90476
A;Molecule type: mRNA
A;Residues: 425-1250,'X',1252-1328,'S',1330-1390,'X',1392-1464 <BRR>
A;Cross-references: GB:K01228; NID:g180391; PIDN:AAA51995.1; PID:g180392
A;Note: sequence partially completed for missing nucleotides by A29439
R;Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.
J. Biol. Chem. 260, 691-694, 1985
A;Title: Multiexon deletion in an osteogenesis imperfecta variant with increased type I
A;Reference number: A22161; MUID:85104931; PMID:2981843
A;Accession: A22161
A;Molecule type: DNA
A;Residues: 472-594,'R',596-607 <CH3>
A;Cross-references: GB:K03178; GB:K03179; NID:g179612; NID:g179611; PIDN:AAA51847.1; PI
A;Note: the authors translated the codon CGT for residue 595 as Pro
R;Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.
Am. J. Hum. Genet. 46, 1034-1040, 1990
A;Title: Variable expression of osteogenesis imperfecta in a nuclear family is explaine

A;Reference number: A35336; MUID:90252792; PMID:2339700
A;Accession: A35336
A;Molecule type: mRNA
A;Residues: 710-720,'E',722-737,'E',739-745 <WAL>
A;Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
R;Forlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Motte
Hum. Mol. Genet. 3, 2201-2206, 1994
A;Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the
A;Reference number: I54365; MUID:95187161; PMID:7781420
A;Accession: I54365
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 746-766,'S',768-781 <FOR>
A;Cross-references: GB:L47667; NID:g1009093; PIDN:AAB59576.1; PID:g1009094
R;Chessler, S.D.; Wallis, G.A.; Byers, P.H.
J. Biol. Chem. 268, 18218-18225, 1993
A;Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of t
A;Reference number: A47426; MUID:93352646; PMID:8349697
A;Accession: A47426
A;Molecule type: mRNA
A;Residues: 1179-1276,'H',1278-1336,1339-1387,'R',1389-1464 <CHE>
A;Cross-references: GB:S64596; NID:g407589; PIDN:AAB27856.1; PID:g407590
A;Note: sequence extracted from NCBI backbone (NCBIN:136444, NCBIP:136445)
A;Note: does not represent an experimentally determined sequence but three different mu
A;Accession: B47426
A;Molecule type: mRNA
A;Residues: 1179-1464 <CH4>
A;Experimental source: normal dermal fibroblast culture
A;Accession: C47426
A;Molecule type: mRNA
A;Residues: 1179-1276,'H',1278-1464 <CH5>
A;Experimental source: fetal cell 86-237
A;Accession: D47426
A;Molecule type: mRNA
A;Residues: 1179-1336,1339-1464 <CH6>
A;Experimental source: fetal cell 86-146
A;Accession: E47426
A;Molecule type: mRNA
A;Residues: 1179-1387,'R',1189-1464 <CH7>
A;Experimental source: fetal cell 88-251
R;Cohn, D.H.; Apone, S.; Eyre, D.R.; Starman, B.J.; Andreassen, P.; Charbonneau, H.; Ni
J. Biol. Chem. 263, 14605-14607, 1988
A;Title: Substitution of Cysteine for Glycine within the Carboxyl-terminal Telopeptide
A;Reference number: I55269; MUID:89008319; PMID:3170557
A;Accession: I55269

Left column

```
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1187-1194, 'C', 1196-1220 <COH>
A;Cross-references: GB:M23213; NID:g340842; PIDN:AAB59363.1; PID:g499622
A;Note: mutant sequence from a patient with mild osteogenesis imperfecta
R;Maekelae, J.K.; Raassina, M.; Virta, A.; Vuorio, E.
Nucleic Acids Res. 16, 349, 1988
A;Title: Human pro-alpha-1(I) collagen: cDNA sequence for the C-propeptide domain.

Query Match      96.6%;    Score 2966;  DB 1; Length 1464;
Best Local Similarity      95.6%;    Pred. No. 5e-137;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

Qy    1 GSSEGPEGVRGEPGPPGPAGAPGADGPGADGEPGAKGADGAPGIAGAPGFPGGARGPSGPE 60
         |||| ||||||||||||||||||| :|| ::|||||||||||||||||||||||||||||:
Db  362 GSSEPQGVRGEPGPPGPAGAAGNPGADGQPGAKGANGAPGIAGAPGFPGGARGPSGPQ 421

Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEGKGRARGEGPGPTGLP 120
         ||||||||||:||||||||||||||||||||||||:|||||||||| |||||||||||||
Db  422 GPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEGKRARGEGPGPTGLP 481

Qy  121 GPPGBRGGPGSRGFPGADGVAGPKGPAGERGSSPPAGPKGSPGEAGRPGEAGLPGAKGLT 180
         |||| |||||||||||||||||||||||||| |||:|||||||||||||||||||||||||
Db  482 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSSPGEAGRPGEAGLPGAKGLT 541

Qy  181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGPGARGEAGVMGFPGPKGAAGEPGKAGERGVP 240
         |||||||||||||||||||:|||||||||||||||:|||||||||||||||||||||||||
Db  542 GSPGSPGPDGKTGPPGPAGQDGRPGPPGPPGPGARGQAGVMGFPGPKGAAGEPGKAGERGVP 601

Qy  241 GPAGVPGAGKDGEAGAEGPPGPAGPAGRGERGEEGPAGSPGFEGLPGPAGPPGEAGKPGEE 300
         ||||||||||||||||||||||||||| |||:|||||||||:||||||||||||||||||:
Db  602 GPAGVPGAGKDGEAGAEGPPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQ 661

Qy  301 GVPGDLGAPGPSGARGEPGPPGPERGVEGPPGPAGPGPPGPAGPRGANGAPGNDGAKGDAGAPGADGSE 360
         ||||||||||||||||| ||| ||||:||||||||||||||| :|||||||||||||||||||||||:
Db  662 GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQ 721

Qy  361 GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
         |||| :|||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db  722 GAPGLQGMPGERGAAGLPGGKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 781

Qy  421 GESGSPGAPGVPTGARGAPGPGRGEPGPPGPAGPAGPGGEPGDAKGEPGDAGAKGDAGPP 480
         ||||||||||||||||||||: |||||||||||||||||||||||||||||||||||||||
Db  782 GESGSPGAPGVPTGARGAPGDRGEPGPPGPAGPAGPAGPGGEPGDAKGEPGDAGAKGDAGPP 841

Qy  481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPGPSGDAGPPGPGPGPA 540
         ||||||||||||||:||||||||||||||||:|||||||||||||||| |||:|||||||||
Db  842 GPAGPAGPPGPIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPGPGPA 901

Qy  541 GKEG 544
         ||||
Db  902 GKEG 905
```

RESULT 2

S21626

collagen alpha 1(I) chain precursor - mouse

```
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 25-Apr-1997 #text_change 13-Aug-1999
C;Accession: S57243; S16374; A23982; I49559; I49557; S39789; I48300; S21626
R;Li, S.W.; Khillan, J.; Prockop, D.J.
Matrix Biol. 14, 593-595, 1994
A;Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I
A;Reference number: S57243
A;Accession: S57243
A;Molecule type: mRNA
A;Residues: 1-1453 <LIS>
A;Cross-references: EMBL:U08020; NID:g470673; PIDN:AAA88912.1; PID:g470674
R;Metsäeranta, M.; Toman, D.; de Crombrugghe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A;Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A;Reference number: S16176; MUID:91274355; PMID:2054384
A;Accession: S16374
```

Right column

```
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1442-1453 <MET>
A;Cross-references: EMBL:X57981; NID:g50484; PIDN:CAA41046.1; PID:g50485
R;French, B.T.; Lee, W.H.; Maul, G.G.
Gene 39, 311-312, 1985
A;Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.
A;Reference number: A23982; MUID:86137403; PMID:3861523
A;Accession: A23982
A;Molecule type: mRNA
A;Residues: 518-1128 <FRE>
A;Cross-references: GB:M14423; NID:g192261; PIDN:AAA37333.1; PID:g192262
R;Monson, J.M.; Friedman, J.; McCarthy, B.J.
Mol. Cell. Biol. 2, 1362-1371, 1982
A;Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for
A;Reference number: I49959; MUID:83141374; PMID:6298597
A;Accession: I49959
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 735-1130 <RES>
A;Cross-references: GB:M17491; NID:g192263; PIDN:AAA37334.1; PID:g192264
R;Harbers, K.; Kuehn, M.; Delius, H.; Jaenisch, R.
Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984
A;Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads
A;Reference number: I49557; MUID:84170331; PMID:6324198
A;Accession: I49557
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-25 <RE2>
A;Cross-references: GB:K01688; NID:g192246; PIDN:AAA37330.1; PID:g553881
R;Fenton, S.P.; Lamande, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.
Biochim. Biophys. Acta 1216, 469-474, 1993
A;Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.
A;Reference number: S39789; MUID:94092741; PMID:8268229
A;Accession: S39789
A;Molecule type: 1-80, 'B', 82-105, 'D', 107-185;1031-1201,'G',1203-1218,'T',1223-
A;Residues: 1-80, 'E', 82-105,'D',107-185;1031-1201,'G',1203-1218,'E',1220-1221,'T',1223-
R;Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.
Mol. Cell. Biol. 14, 5950-5960, 1994
A;Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indire
A;Reference number: I48300; MUID:94344105; PMID:8065328
A;Accession: I48300
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-80, 'B', 82-105,'D', 107-147 <REF>
A;Cross-references: EMBL:X54876; NID:g50486; PIDN:CAA38657.1; PID:g50487
C;Genetics:
A;Gene: COL1A1
A;Introns: 770/3; 788/3; 806/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 1058
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: coiled coil; extracellular matrix; glycoprotein; heterotrimer; triple helix
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-151/Domain: amino-terminal propeptide #status predicted <PRO>
F;30-89/Domain: von Willebrand factor type C repeat homology <VWC>
P;152-1453/Product: collagen alpha 1(I) chain #status predicted <MAT>
P;1224-1453/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match      93.1%;    Score 2857;  DB 2; Length 1453;
Best Local Similarity      91.4%;    Pred. No. 9.5e-132;
Matches 497; Conservative 28; Mismatches 19; Indels 0; Gaps 0;

Qy    1 GSSEGPEGVRGEPGPPGPAGAPGADGADGPGADGEPGAKGADGAPGIAGAPGFPGGARGPSGPE 60
         ||||| |:|||||||||||||||||||| ||||||||:|||||||::|||||||||||||||
Db  351 GSSEPQGVRGEPGPPGPAGAAGNPGADGQPGAKGANGADGQPGAKGANGADGPGAKGANGAPGFPGARGPGFPGARGPSGPQ 410

Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGEGPPGPAGEGKGRARGEGPGPTGLP 120
         ||||||||||:||||||||||||||||||||| |||:||:||:|||||||||
Db  411 GPSGPPGPKGNSGEPGAPGNKGDTGAKGEPGATGVQGPPGPAGEEGKRARGEGPGPSCLP 470

Qy  121 GPPGBRGGPGSRGFPGADGVAGPKGPGAGERGSSPPAGPKGSPGEAGRPGEAGLPGAKGLT 180
         |||||||||||:|||||||||||||||||||| || ||| |||||:|||||||||||||||
Db  471 GPPGERGGPGSRGFPGADGVAGPKGPGSGERGAPGAPGPAGPKGSPGEAGRPGEAGLPGAKGLT 530
```

Qy 181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGPARGEAGVMGFPGPKGAAGEPGKAGERGVP 240
Db 531 GSPGSPGPDGKTGPPGPAGQDGRPGPAGPPGPARGQAGVMGFPGPKGTAGEPGKAGERGLP 590

Qy 241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEEGAQGPAGPPGPAGEGGPAG 300
Db 591 GPPGAVGPAGKDGEAGAQGPAGPAGEGPAG 650

Qy 301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGPGADGAPGDDGAKDAGAPGAPGSE 360
Db 651 GVPGDIGAPGPSGARGERGFPGERGVQGPPGPAGPRGNNGAPGNDGAKGDTGAPGAPGSQ 710

Qy 361 GAPGLEGMPGERGAAGLPGPKGDRDGAPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
Db 711 GAPGLQGMPGERGAAGLIPGPKGDRDGAPKGADGSPGKDGARGLTGPIGPPGPAGAPGDK 770

Qy 421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGERGEPGDAGEPGDTGVKGDAGPP 480
Db 771 GEAGPSGPPGPTGARGAPGDRGEAGPPGPAGPPGPAGDGQPGAKGEPGDTGVKGDAGPP 830

Qy 481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGPAAGRVGPPGPGDGAPGPPGPA 540
Db 831 GPAGPAGPPGPIGNVGAPGPKGPRGAAGPPGPAAGRVGPPGPSGNAGPPGPPGPV 890

Qy 541 GKEG 544
Db 891 GKEG 894

RESULT 3
CGCHlS
collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)
C;Species: Gallus gallus (chicken)
C;Date: 12-Aug-1981 #sequence_revision 06-Jul-1982 #text_change 31-Mar-2000
C;Accession: A90458; A90181; A02857
R;Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Kang, A.H.; Gross, J.
Biochemistry, 21, 2048-2055, 1982
A;Title: Amino acid sequence of chick skin collagen alpha1(I)-CB8 and the complete prima
A;Reference number: A90458; MUID:82231995; PMID:7093229
A;Accession: A90458
A;Molecule type: protein
A;Residues: 1-1036 <HIG>
A;Experimental source: skin
A;Note: this is the latest in a series of papers from these workers elucidating the sequ
R;Eyre, D.R.; Glimcher, M.J.
Biochem. Biophys. Res. Commun. 48, 720-726, 1972
A;Title: Evidence for a previously undetected sequence at the carboxyterminus of the alp
A;Reference number: A90181; MUID:72243016; PMID:5047697
A;Accession: A90181
A;Molecule type: protein
A;Residues: 1037-1042 <EYR>
A;Experimental source: skin
A;Note: residues 1037-1042 above correspond to the carboxyl end of the protein
C;Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some
C;Comment: Most of the prolines at the third position of the tripeptide repeating unit (
C;Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in po
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
A;Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trimer;
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match   90.4%; Score 2775; DB 1; Length 1042;
Best Local Similarity 89.0%; Pred. No. 6.9e-128;
Matches 484; Conservative 28; Mismatches 32; Indels 0; Gaps 0;

Qy   1 GSEGPEGVRGEPGPPGPAGAAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE 60
Db 200 GSSEGPQGSRGEPGPPGPAGAAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPQ 259

Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEEGKPGARGEPGPTGLP 120
Db 260 GPSGAPGPKGNSGEPGAPGNKGDTGAKGEPGPAGVQGPPGPAGEEGKRGARGEPGPAGLP 319

Qy 121 GPPGERGGPGGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180

Db 320 GPAGERGAPGASRGPPGADGIAGPKGPPGERGSPGAVGPKGSPGEAGRPGEAGLPGAKGLT 379
Qy 181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGPARGEAGVMGFPGPKGAAGEPGKAGERGVP 240
Db 380 GSPGSPGPDGKTGPPGPAGQDGRPGPAGPPGPARGQAGVMGFPGPKGPAGEPGKPGERGAP 439
Qy 241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEGAPGSPGFEGLPGPAGPKGPGEAGKPGEE 300
Db 440 GPPGAVGAAGKDGEAGAQGPPGPGPTGPGPAGPFQGLPGPAGPPGPEAGKPGEQ 499
Qy 301 GVPGDLGAPGPSGARGEGFPGERGVEGPPGPAGPPGPADGAPGDGDKAGDAPGAPGSE 360
Db 500 GVPGNAGAPGPAGARGERGFPGERGVQGPPGPQGPRGANGAPGNDGAKGDAGAPGAPGNE 559
Qy 361 GAPGLEGMPGERGAAGLPGPKGDRDGAPGKGDGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
Db 560 GPPGLEGMPGERGAAGLPGAKGDRGDPGPKGADGAPGKDGLRGLTGPIGPPGPAPGDK 619
Qy 421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPADGEPGDAKGEPGDAGAKGDAGPP 480
Db 620 GEAGPPGPAGPTGARGAPGDRGEPGPPGPAGPFAGPGAGDGPGAKGETGDAGAKGDAGPP 679
Qy 481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGDATGFPGAAGRVGPPGPSGDAGPPGPPGPA 540
Db 680 GPAGPTGAPGPAGZVGAPGPKGARGSAGPPGDATGFPGAAGRVGPPGPSGNIGLPGPPGPA 739

Qy 541 GKEG 544
Db 740 GKZG 743

RESULT 4
T45467
collagen alpha 1(II) chain precursor [imported] - horse
N;Alternate names: type II collagen
C;Species: Equus caballus (domestic horse)
C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 04-Mar-2000
C;Accession: T45467
R;Richardson, D.W.; Dodge, G.R.
submitted to the EMBL Data Library, June 1996
A;Description: Cloning of equine type II collagen and modulation of its expression in e
A;Reference number: Z22977
A;Accession: T45467
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1418 <RIC>
A;Cross-references: EMBL:U62528; PIDN:AAB05773.1
C;Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology

Query Match   72.3%; Score 2220; DB 2; Length 1418;
Best Local Similarity 72.6%; Pred. No. 6.5e-101;
Matches 395; Conservative 37; Mismatches 112; Indels 0; Gaps 0;

Qy   1 GSEGPEGVRGEPGPPGPAGAAGDPGADGPDGADGAPGDAGEPGAGAPGPAGAGAGAPGSGPE 60
Db 315 GPBEAGQPGRGEPGTPGSSPGPAGAAGNPGDTGIPGAKGSAGAPGIAGAPGFPGPRGPPGPQ 374
Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEGKPGARGEPGPTGLP 120
Db 375 GATGPLGPKGQTGEPGIAGFKGEQGPKGEPGPAGERGSPGAGERGSPGEAGRPGEAGPGAPV 434
Qy 121 GPPGERGGPGGSRGFPGADGVAGPKGPAGERGSPGEAGRPGEAGLPGAKGLT 180
Db 435 GPPGERGAPGNRGFPGGQDGLAGPKGAPGERGPPGPAGEAGVMGFPGPKGAANGDPGRPGEGPGLPGARGLT 494
Qy 181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGPARGEAGVMGFPGPKGAAGEPGKAGERGVP 240
Db 495 GRPDAGPQGKVGPSGAPGEDGRPGPPGPGPAGERGEEGAPGSPGFEGLPGPAGEEKGLP 554
Qy 241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEEGAPGSPGFEGLPGPAGPPGPEAGKPGEE 300
Db 555 GAPGLRGLPKGDGETGAAGPPGPAGPAGERGEQGAPGAGERGEQGAPPGPGDQ 614

```
Qy  301  GVPGDLGAPGPSGARGEPGPGPGERGVEGPPGPAGPPGADGAPGDDGAKGDAGAPGPGSE  360
              |||: |||: ||| ::|||||| |||||||  |||| ||| :|||| :||::
Db  615  GVPGEAGAPGLVGPRGERGPPGERGSPGAQGLQGARGLPGTPGTDGPKGASGPAGPPGAQ  674

Qy  361  GAPGLESMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  420
              ||:||| :|||||||||:|:| :|| |||||||||| ||||:|||:|| |||| |||:||
Db  675  GPPGLQGMPGERGAAGIAGPKGDRGDVGEKGPEGAPGKDGDRGLTGPIGPPGPPGPAGANGEK  734

Qy  421  GESGSPGSPAGPTGARGAPGDRGEPGEPGPPGPAGPPGADGEPGAKGEPGDGAKGDAGPP  480
              :|| |||| ||||::||||| ||| |||:::||||||||||| :|||||| |||||||
Db  735  GEVGPPGPAGTAGARGAPGERGETGPPGPAGPAGPPGADGDQGPGAKGEQGEAGQKGDAGAP  794

Qy  481  GPAGPAGPPGPIGDVGAPGAKGARGAGSAGPPGATGFPGAAGRVGPPGPSGDAGPPGPPGPA  540
              ||| ||||:|:| ||||| ||||::||| |||:::||||||||||| :|| ||::
Db  795  GPQGPSGAPGPGQPTGVTGPKGARGAQGPPGATGFPGAAGRVGPPGGSNGNPGPPGPPGPS  854

Qy  541  GKEG  544
              ||:|
Db  855  GKDG  858
```

RESULT 5

CGHU6C

collagen alpha 1(II) chain precursor [validated] - human

N;Alternate names: procollagen alpha 1(II) chain
N;Contains: chondrocalcin; collagen alpha 1(II) chain precursor splice form 1; collagen
C;Species: Homo sapiens (man)
C;Date: 28-May-1986 #sequence revision 01-Sep-1995 #text change 08-Dec-2000
C;Accession: A38513; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S64674; S63
7250; I37251; I37252; I37253; I37254; I55338; I59535; I61910
R;Ryan, M.C.; Sieraski, M.; Sandell, L.J.
Genomics 8, 41-48, 1990
A;Title: The human type II procollagen gene: identification of an additional protein-cod
A;Reference number: A38513; MUID:91184811; PMID:2081599
A;Accession: A38513
A;Molecule type: DNA
A;Residues: 1-103 <RYA>
A;Cross-references: EMBL:X16711; NID:g30040; PIDN:CAA34683.1; PID:g30041
A;Note: alternative splice form 1
R;Ryan, M.C.; Sandell, L.J.
J. Biol. Chem. 265, 10334-10339, 1990
A;Title: Differential expression of a cysteine-rich domain in the amino-terminal propep
A;Reference number: A35428; MUID:90285153; PMID:2355003
A;Accession: A35428
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 27-81,'L',83-103 <RYA2>
A;Note: alternative splice form 2; splicing appears to be under developmental regulatio
R;Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.
Genomics 4, 438-441, 1989
A;Title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide con
A;Reference number: A30147; MUID:89233138; PMID:2714801
A;Accession: A30147
A;Molecule type: DNA
A;Residues: 104-157,'P',159-236 <SUM>
A;Cross-references: GB:J03065; GB:M23660; GB:M25655; GB:M25656; GB:M25730; GB:M32168; G
R;Ala-Kokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.
Proc. Natl. Acad. Sci. U.S.A. 87, 6565-6568, 1990
A;Title: Single base mutation in the type II procollagen gene (COL2A1) as a cause of pr
A;Reference number: A94227; MUID:90370826; PMID:1975693
A;Accession: A33116
A;Molecule type: DNA
A;Residues: 171-172,'C',174-175 <ALA>
A;Note: mutant sequence from a family with family with primary generalized osteoarthrit
R;Diab, M.; Wu, J.J.; Eyre, D.R.
Biochem. J. 314, 327-332, 1996
A;Title: Collagen type IX from human cartilage: a structural profile of intermolecular
A;Reference number: S64673; MUID:96195147; PMID:8660302
A;Accession: S64674
A;Molecule type: protein
A;Residues: 188-189,'X',191-195;1224-1230,'X',1232-1236 <DIA>
R;Franc, S.; Marzin, E.; Boutillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage,
Eur. J. Biochem. 234, 125-131, 1995
A;Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil ca
A;Reference number: S63514; MUID:96096730; PMID:8529631
A;Accession: S63514
A;Molecule type: protein
A;Residues: 243-261;575-590;756-763,'X',765-779 <FRA>
R;Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.
R;Tiller, G.E.; Weis, M.A.; Polumbo, P.A.; Gruber, H.E.; Rimoin, D.L.; Eyre.
Nucleic Acids Res. 17, 9473, 1989
Am. J. Hum. Genet. 56, 388-395, 1995
A;Title: Nucleotide sequence of the full length cDNA encoding for human type II procolla
A;Title: An RNA-splicing mutation (G+5IVS20) in the type II collagen gene.
A;Reference number: S06715; MUID:90067946; PMID:2587267
A;Reference number: I38867; MUID:95150028; PMID:7847372
A;Accession: S06715
A;Accession: I38867
A;Molecule type: mRNA
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Residues: 1-28,'R',99-1487 <SU2>
A;Molecule type: DNA
A;Cross-references: EMBL:X16468; NID:g29515; PIDN:CAA34488.1; PID:g29516
A;Residues: 440,'G',442-456,'E',458-480,'P',482-509 <TIl1>
A;Note: alternative splice form 1
A;Cross-references: EMBL:U15199; NID:g557053; PIDN:AAB60370.1; PID:g557054
R;Vikkula, M.; Metsaeranta, M.; Syvaenen, A.C.; Ala-Kokko, L.; Vuorio, E.; Peltonen, L.
R;Ramirez, F.
Biochem. J. 285, 287-294, 1992
submitted to the EMBL Data Library, December 1988
A;Title: Structural analysis of the regulatory elements of the type-II procollagen gene.
A;Reference number: S24270; MUID:92344585; PMID:1637314
A;Accession: S04892
A;Accession: S24270
A;Molecule type: mRNA
A;Status: translation not shown
A;Residues: 501-676,'A',678-783,'A',785-831,'PA',834,'F',836-1214 <RAM>
A;Molecule type: DNA
A;Cross-references: EMBL:X13783; NID:g30037; PIDN:CAA32030.1; PID:g930050
A;Residues: 1-28 <VIK>
R;Vikkula, M.; Peltonen, L.
A;Cross-references: EMBL:X58709; GB:S40537; NID:g35659
FEBS Lett. 250, 171-174, 1989
A;Note: this translation is not annotated in GenBank entry HSPROCOE1, release 111.0
A;Title: Structural analyses of the polymorphic area in type II collagen gene.
R;Nunez, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.
A;Reference number: S05000; MUID:89325561; PMID:2753125
Gene 44, 11-16, 1986
A;Accession: S05000
A;Title: Promoter region of the human pro-alpha-1-(II)-collagen gene.
A;Molecule type: DNA
A;Reference number: A24828; MUID:87031574; PMID:3021582
A;Residues: 630-640,'A',642-785 <VIK2>
A;Accession: A24828
A;Cross-references: EMBL:X16158; NID:g29951; PIDN:CAA34278.1; PID:g1335018; PIDN:CAA342'
A;Molecule type: DNA
PIDN:CAA34283.1; PID:g1335023; PIDN:CAA34284.1; PID:g1335024
A;Residues: 1-8,'T',10-28 <NUN>
R;Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, '
A;Cross-references: GB:M25698; NID:g180872; PIDN:AAA52051.1; PID:g553237
J. Biol. Chem. 267, 22522-22526, 1992
R;Baldwin, C.T.; Reginato, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.
A;Title: An amino acid substitution (Gly853-->Glu) in the collagen alpha 1(II) chain pr
Biochem. J. 262, 521-528, 1989
A;Reference number: A44309; MUID:93054548; PMID:1429602
A;Title: Structure of cDNA clones coding for human type II procollagen. The alpha-1(II)
A;Accession: A44309
A;Reference number: S06496; MUID:90026318; PMID:2803268
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Accession: S06496
A;Molecule type: DNA; mRNA
A;Molecule type: mRNA
A;Residues: 7-28,'R',99-157,'P',159-440,'G',442-456,'E',458-640,'A',642-831,'PA',834,'F'

A;Residues: 752-831,'PA',834,'F',836-1005,'K',1007-1036,'Q',1038-1052,'E',1054-1068,'T',
A;Cross-reference: GB:L00977; NID:g180811; PIDN:AAB23914.1; PID:g258774
A;Note: sequence extracted from NCBI backbone (NCBIP:117273); Parts of this sequence wer
A;Note: this translation is not annotated and this publication is not cited in GenBank e
A;Note: mutant sequence associated with perinatal lethal hypochondrogenesis
R;Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.
Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990
A;Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individua
A;Reference number: S16502; MUID:90251662; PMID:2339128
A;Accession: S16502
A;Molecule type: DNA
A;Residues: 1164-1184,'GPSGKDGANGIPGPI',1185-1199 <TIL2>
A;Cross-references: EMBL:M37126; NID:g180808; PIDN:AAA52037.1; PID:g180809
A;Note: mutant sequence from a patient with spondyloepiphyseal dysplasia
R;Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.
Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985
A;Title: Identification and characterization of the human type II collagen gene (COL2A1)
A;Reference number: A02858; MUID:85190534; PMID:3857598
A;Accession: A02858
A;Molecule type: DNA
A;Residues: 1032-1056,'N',1058-1068,'T',1070-1487 <CHE>
A;Cross-references: GB:J00116; NID:g180395; PIDN:AAA51997.1; PID:g180396
R;Elima, K.; Vuorio, T.; Vuorio, E.
Nucleic Acids Res. 15, 9499-9504, 1987
A;Title: Determination of the single polyadenylation site of the human pro-alpha-1(II) g
A;Reference number: A27280; MUID:88067771; PMID:2825137
A;Accession: A27280
A;Molecule type: DNA; mRNA
A;Residues: 1175-1487 <ELI>
A;Cross-references: EMBL:X06268; NID:g30096; PIDN:CAA29604.1; PID:g30097
A;Experimental source: fetal epiphyseal cartilage
R;van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
Biochem. J. 237, 923-925, 1986
A;Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
A;Reference number: A57033; MUID:87099927; PMID:3800925
A;Accession: A57033
A;Molecule type: protein
A;Residues: 'XE',1244-1246,'N',1248,'X',1250-1265;1295-1305;1395-1408 <VAN>
A;Note: chondrocalcin identified as released collagen 1(II) chain carboxyl-terminal prop
R;Strom, C.M.; Upholt, W.B.
Nucleic Acids Res. 12, 1025-1038, 1984
A;Title: Isolation and characterization of genomic clones corresponding to the human typ
A;Reference number: A21733; MUID:84118798; PMID:6320112
A;Accession: A21733
A;Molecule type: DNA
A;Residues: 1245-1295 <STR1>
A;Cross-references: EMBL:X00339; EMBL:X00298; NID:g394699; PIDN:CAA25092.1; PID:g4378975
A;Accession: B21733
A;Molecule type: DNA
A;Residues: 894-909,'PE' <STR2>
A;Cross-references: GB:K01785; NID:g30035; PIDN:CAA25082.1; PID:g1335032
R;Nunez, A.M.; Francomano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
Biochemistry 24, 6343-6348, 1985
A;Title: Isolation and partial characterization of genomic clones coding for a human pro
gene.
A;Reference number: A24561; MUID:86104139; PMID:3002437
A;Accession: A24561
A;Molecule type: DNA
A;Residues: 1296-1358 <NUN2>
A;Cross-references: GB:M12048; NID:g180017
A;Note: this translation is not annotated in GenBank entry HUMCCT2A, release 111.0
A;Note: the codons given for 1333-Lys (AGG) and 1350-Gly (GCA) are inconsistent with the
R;Sangiorgi, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramirez,
Nucleic Acids Res. 13, 2207-2225, 1985
A;Title: Isolation and partial characterization of the entire human pro alpha 1(II) coll
A;Reference number: 137249; MUID:85215609; PMID:2987845
A;Accession: S59491
A;Molecule type: DNA
A;Residues: 1-1419 <MET>
A;Cross-references: GB:M65161
R;Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.
Development 111, 945-953, 1991
A;Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartila
A;Reference number: A44885; MUID:91347939; PMID:1879363
A;Molecule type: DNA
A;Residues: 7-28 <SAN2>

A;Cross-references: GB:M23759; NID:g180845; EMBL:X03320; GB:M24938; NID:g30104
A;Note: the GenBank PID is based on an incorrect reading frame
A;Accession: 137250
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 541-560 <SAN3>
A;Cross-references: EMBL:X02378; GB:M23870; NID:g30107; PIDN:CAA26227.1; PID:g929621
A;Accession: 137251

Query Match   71.7%; Score 2202; DB 1; Length 1487;
Best Local Similarity   71.9%; Pred. No. 5e-100;
Matches 391; Conservative 40; Mismatches 113; Indels 0; Gaps 0;

Qy   1 GSEGPEGVRGPPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAGPPGARGPSGPE 60
         ||||| :|| ||:|| | | |||| ||| | | | | | | ||:|
Db 384 GPEGAQGPRGEPGTPGSPGPAGASGNPGTDGIPGAKGSAGAPGIAGAGPFGPGPRGPPDPQ 443

Qy 61 GPGGPPGPGKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGENGKPGARGEPGPTGLP 120
         ||| | ||:||::: |:| ||||
Db 444 GATGPLGPKGGQTGKPGIAGFKGEQGPKGPAGERGRSPGPAGPKGSPGEARGRGAGEPGGVGPI 503

Qy 121 GPPGERGGPGSRGFPGADGVAPGKPGAGERGSPGPAGPKGSPGEARGRPGEAGLPGAKGLT 180
         ||||| :||:|| | :||||| | |:|| ||||| |
Db 504 GPPGERGAPGNRGFPGQDGLAGPKGAPGERGPSGLAGPKGANGDPGRPGEPGLPGARGLT 563

Qy 181 GSPGSPGPDGKTGPPGPAGEDGRPGEPGPPGPPGAGERGEGPAGSPGFEGLPGPAGPPGEE 240
         | |||||| ||| | | | | | | || | | | || | | | | | |
Db 564 GRPGDAGPQGKVGPSGAPGEDGRPGPPGPGPQQCARGQPGVMGFPGPAGKGKAGEKGLP 623

Qy 241 GPPGAVGPAGKDGEAGAEGPPGPPGPAGPAGERGEGPAGSPGFEGLPGPAGPPGEAGKPGEE 300
         | || ||||:|| | ||||| || | | | || || | | | | | ||:||
Db 624 GAPGLRGLPGKDGETGAEGPPGPAGEDGPAGERGEQGARGEGQAGPSGPGPGPPGEGGKPGDQ 683

Qy 301 GVPGDLGAPGPSGCARGEPGFPGERGVEGPPGPAGPAGDGAKGDAGAPGAPGSE 360
         | || ||||| | | || | | |||| || | | | | | | |
Db 684 GVPGEAGAPGLVGPRGERGFPGERGSPGAQGLQGPRGLPGTPGTDGPKGASGPAGPPGAQ 743

Qy 361 GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
         | || ||||||| |:|: | | | | | | || | | | | | | | | |:
Db 744 GPPGLQGMPGERGAAGIAGPKGDRGDVGEKGPEGAPGKDGGRGLTGPIGPFGPAGANGEK 803

Qy 421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPGDAGAKGDAGAPGAGKPGDAGAKGDAGPP 480
         | | |||||| | | | | | |||| || || | | | | | | | | |
Db 804 GBVGPPGPAGSAGARGARGERGETGPPGTSGIAGPPGADGQPGAKGEQGEAGQGKGDAGAP 863

Qy 481 GPAGPAGPPGIGDVGAPGAKGARGSAGPPGATGFEGAAGRVGPPGSGDAGPPGPPGPA 540
         | || ||| | | | | | | |: | : | | | || | | | | |||| |
Db 864 GPQGPSGAPGPQGPTGVTGPKGARGAQGPPGATGFPGAAGRVGPPGSNGNPGGPPGPS 923

Qy 541 GKEG 544
         |:|
Db 924 GKDG 927

RESULT 6
A41182
collagen alpha 1(II) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
C;Accession: A41182; A44885
R;Metsaeranta, M.; Toman, D.; de Crombrugghe, B.; Vuorio, E.
J. Biol. Chem. 266, 16862-16869, 1991
A;Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
A;Reference number: A41182; MUID:91358489; PMID:1885613
A;Accession: A41182
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-1419 <MET>

A;Accession: A44885
A;Molecule type: DNA
A;Residues: 1-28 <CHE>
A;Cross-references: GB:S63190; NID:g234368; PIDN:AAB19627.1; PID:g234369
A;Note: sequence extracted from NCBI backbone (NCBIN:63190, NCBIP:63192)
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trime
F;191-1419/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match   71.3%;   Score 2189; DB 2;   Length 1419;
Best Local Similarity   71.7%;   Pred. No. 2.1e-99;
Matches 390; Conservative 39; Mismatches 115;   Indels 0;   Gaps 0;

Qy   1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGPGARGPSGPE 60
           ||||||||||||||||||| :||||| :||| :||:||||||||||||||||:
Db   316 GPEGAQGSRGEPGNPGSPGPAGASGNPGTDGIPGAKGSAGAPGIAGAPGPGPRGPGPPGPQ 375

Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPVGVEGPPGKGARGEPGPTGLP 120
           ||||||||||||||||||||| :|||||||||||||||||| :||||||:||||:
Db   376 GATGPLGPKGQAGEPGIAGFKGDQGPKGETGPAGPQGAGEPGAGEGKRGARGEPGGAGPI 435

Qy   121 GPGERGRGGSRGFPGADGVAGPKGAGERGSPGPAGERGSPGPAGRPGEAGLPGAKGLT 180
           |||||||||||||| |||||||:||||| |||||||||||:||||||||||||||||
Db   436 GPPGERGAPGNRGFPGQDGLAGPKGAGERGPSGLAGPKGANGDPGRPGEPGLPGARGLT 495

Qy   181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPGPGARGEAGVMGFPGPKGAAGEPGKAGERGVP 240
           ||||||||||| :||||||||||||||| |||||||||||||||| |||||||:
Db   496 GRPGDAGPQGKVGPSGAPGEDGRPGPPGPGPGARGQPGVMGFPGPKGANGDPGVKGKAGEKGLA 555

Qy   241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEEGPAGSPGFEGLPGPAGPAGEGKPGEE 300
           |||||||||||||||| :||||||||||| |||||||||| |||||||||||||||||
Db   556 GAPGLRGLPGKDGETGAAGEPGPSGSPGAGERGEQGAPGPSGFQGLPGPGPAGEKGLA 615

Qy   301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGADGAKGDADAGAPGAPGSE 360
           :|||:||||||||||||||:||| ||||||||||||||||:|||||:|||||||||:
Db   616 GIPGEAGAPGLVGPRGERGFPGERGSPGAQLQGPRGLPGTPGTDGPKGAAGPDGPPGAQ 675

Qy   361 GAPGLEGMPGERGAAGLPGPKGDRDGAGPSGPKGDGVRGLTGPIGPPGPAGAPGDK 420
           |||||||||||||||| :|||||||||||||||||||| :|||||||||||:
Db   676 GPPGLQGMPGERGAAGIAGPKGDRGDVGEKGPEGAPGKDGGRGLTGPIGPPGPAGANGEK 735

Qy   421 GESGPSGPAGPTGARGAPGDGRGEPGPPGPAGPAGPPGCATGFPGAAGRVGPPGPSGDAGPPGPGPA 480
           ||||||||||||||||:|| ||||:|||||||||||:||:||||||:||||||||||||||||
Db   736 GEVGPPGPSSSTGARGAPGEPGETGPPGPAGPAGPPGADGQPGPAGDGRGLTGPIGPPGPAGANGEK 795

Qy   481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGCATGPGAAGRVGPPGPGSGDAGPPGPGPA 540
           ||||||||||||:|| :|| ||:|:||||||||: :|||||||||:|||||||||||
Db   796 GPQGPSGAPGPQGPFTGPVGPKGPKGARGAQGPPGPGATGFPGAAGRVGPPGANGNPGPAGPPGPA 855

Qy   541 GKEG 544
           ||:|
Db   856 GKDG 859

RESULT 7
B41182
collagen alpha 1(II) chain precursor (long splice form) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 16-Jul-1999
C;Accession: B41182
R;Metsaeranta, M.; Toman, D.; de Crombrugghe, B.; Vuorio, E.
J. Biol. Chem. 266, 16862-16869, 1991
A;Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
A;Reference number: A41182; MUID:91358489; PMID:1885613
A;Accession: B41182
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-1487 <MET>
A;Cross-references: GB:M65161
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trime
F;33-91/Domain: von Willebrand factor type C repeat homology <VWC>
F;1259-1487/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match   71.3%;   Score 2189; DB 2;   Length 1487;
Best Local Similarity   71.7%;   Pred. No. 2.1e-99;
Matches 390; Conservative 39; Mismatches 115;   Indels 0;   Gaps 0;

Qy   1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGPGARGPSGPE 60
           ||||||||||||||||||| :||||| :||| :||:|||||||||||||||| :
Db   384 GPEGAQGSRGEPGNPGSPGPAGASGNPGTDGIPGAKGSAGAPGIAGAPGFPGPGRGPPGPQ 443

Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGBEGKPGPGARGEPGPTGLP 120
           ||||||||||||||||||||| :|||||||||||||||||:|| :||:|||||||||:
Db   444 GATGPLGPKGQAGEPGIAGFKGDQGPKGETGPAGPQGAPGAGEEGKRGARGEPGGAGPI 503

Qy   121 GPGERGGRGGSRGFPGADGVAGPKGPAGERGSPGPAGRGSPGEAGLPGAKGLT 180
           |||||| ||||:||||| |||||||:||||| |||||||||:||:|||||||||
Db   504 GPPGGERGAPGNRGFPGQDGLAGPKGAPGERGPSGLAGPKGANGDPGRPGEPGLPGARGLT 563

Qy   181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPGPGARGEAGVMGFPGPKGAAGEPGKAGERGVP 240
           ||||||||||| :||||||||||||||||| |||||||||||||||| ||||||||:
Db   564 GRPGDAGPQGKVGPSGAPGEDGRPGPPGPGPGPQGARGPPGVMGFPGPKGANGEPGKAGEKGLA 623

Qy   241 GPGAVGPAGKDGEAGAEGPPGPGPAGERGEEGPAGSPGFEGLPGPAGPGPGEAGKPGSE 300
           ||||||||||:|||||||||||||||| |||||||||| ||||||||||| |||||||::
Db   624 GAPGLRGLPGKDGETGAAGPGPSGSPPAGERGEQGAPGPSGFQGLPGPGPGEGGKQGDQ 683

Qy   301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPGPAGPPGADGAKGDADAGAPGAPGSE 360
           :|||:|||||||||||||||:||| ||||||||||||||||:|||||:|||||||||::
Db   684 GIPGEAGAPGLVGPRGERGFPGERGSPGAQLQGPRGLPGTPGTDGPKGAAGPDGPPGAQ 743

Qy   361 GAPGLEGMPGERGAAGLPGPKGDRDGAGPSGPKGDGVRGLTGPIGPPGPAGAPGDK 420
           |||||||||||||||| :|||||||||||||||||||| :||||||||||:
Db   744 GPPGLQGMPGERGAAGIAGPKGDRGDVGEKGPEGAPGKDGGRGLTGPIGPPGPAGANGEK 803

Qy   421 GESGPSGPAGPTGARGAPGDGRGEPGPPGPAGPAGPPGCATGFPGAAGRVGPPGPSGDAGPP 480
           ||||||||||||||||:|| ||||:|||||||||||:||:||||||:|||||||||||||
Db   804 GEVGPPGPSSSTGARGAPGEPGETGPPGPAGPAGPPGADGQPGPAGDQGEAGQKGDAGAP 863

Qy   481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGCATGPPGAAGRVGPPGPGSGDAGPPGPGPPA 540
           |||||||||||:|| :|| ||:|:||||||||: :|||||||:|||:||||||||||||
Db   864 GPQGPSGAPGPQGPGTGPVTGPKGARGAQGPPGPGATGFPGAAGRVGPPGANGNPGPAGPPGPA 923

Qy   541 GKEG 544
           ||:|
Db   924 GKDG 927

RESULT 8
CGRTIS
collagen alpha 1(I) chain - rat (tentative sequence) (fragments)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text change 31-Mar-2000
C;Accession: A90559; A90552; A92029; A90353; A90566; A90557; A90362; A90379; A91209; A9
R;Bornstein, P.
Biochemistry 8, 63-71, 1969
A;Title: Comparative sequence studies of rat skin and tendon collagen. II. The absence
A;Reference number: A90559; MUID:69155173; PMID:577344
A;Contents: CNBr0 and CNBr1
A;Accession: A90559
A;Molecule type: protein
A;Residues: 1-19 <BO1>
A;Experimental source: tendon
A;Note: sequences from skin and tendon appear to be identical
A;Note: the amino-terminal tetrapeptide may be removed by limited proteolysis during ex
R;Kang, A.H.; Bornstein, P.; Piez, K.A.
Biochemistry 6, 788-795, 1967
A;Title: The amino acid sequence of peptides from the cross-linking region of rat skin
A;Reference number: A90552; MUID:67162268; PMID:5337086
A;Contents: CNBr1
A;Accession: A90552
A;Molecule type: protein
A;Residues: 5-19 <KAN>
A;Experimental source: skin
R;Bornstein, P.

A;Note: the composition of peptides comprising residues 1-9 and 1-19 confirms the seque
A;Note: this region (residues 651-671 above) probably corresponds to positions 1032-105
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C;Comment: The order of the nine CNBr peptides in the alpha 1(I) chain of rat skin coll
C;Comment: The complete chain contains 1052 residues.
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydroxy
F;1/Modified site: blocked amino end (Glx) (probably pyrrolidone carboxylic acid) #stat
F;9/Modified site: allysine (Lys) #status experimental
F;103,424,547/Binding site: carbohydrate (Lys) (covalent) #status experimental
F;103/Modified site: 5-hydroxylysine (Lys) #status experimental
F;424,547/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental

Query Match   70.2%; Score 2154; DB 1; Length 671;
Best Local Similarity   77.3%; Pred. No. 5.9e-98;
Matches 381; Conservative   28; Mismatches 36; Indels 48; Gaps   2;

Qy   1 GSEGPEGVRGEPPGPPGPAGAAGPAGPDCADGEPGAKGADGADGAGPGARGPSGPE 60
         ||||| ||||||| ||||||| | ||: ||||||||:||||||| ||||||||:
Db 200 GSEGPQGVRGEPGPPGPAGANGPGADGQPGAKGANGAPGIAGAPGIAGAPGPSGPQ 259

Qy   61 GPGGPPPGKGDGGEPGAPGSKDTGAKGEPGVGVEGPPGPAGEEGKPGARGEPGPTGLP 120
         ||.|||.|| |.|.||||| |.|||||. ||: ||||||||.|||.|||||||:|||
Db 260 GPSGAPGPGKGNSGPGAPGNKGDTGAKEPGPAGVQGPPGPAGEBGKRGARGEPGPSGLP 319

Qy 121 GPPGERGPGSRGFPGADGVAGPKGPAGERGPSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
         ||||| ||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db 320 GPPGBERGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 379

Qy 181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPGARGEAGVMGPPGPKGAAGEPGKAGERGVP 240
         |||||||||||||||||||| ||||| |||||| |||||||||||| |||||||||||
Db 380 GSPGSPGPDGKTGPPGPAGZBGRPGPGPGARGQAGVMGPPGPKGTAGEPGKAGERGVP 439

Qy 241 GPPGAVGPAGKDGEAGAEGPPGPAGAGERGEGPAGSPGFEGLPGPAGPPGEAGKPGEE 300
         ||||||||||||||||| |.|||| ||||| ||||||||:|||||||||||||||||
Db 440 GPPGAVGPAGKDGEAGAQGAPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGZZ 499

Qy 301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGADGAKGDDGAGAPGAPGAPGE 360
         |||||||||||||||| |||||||||||||||||| :.|.|..||:|||:||||||||
Db 500 GVPGDLGAPGPSGARGEGRGFPGERGVQGPPGPAGPRGNNGAPGBBGAKGDTGAPGAPGSQ 559

Qy 361 GAPGLEGMPGERGAAGLPGPKGKDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
         |||||.|.|||| ||:|:| GPPGPPGSPGZZ 583
Db 560 GAPGLZGMSGLZ-------------------------GPPGPPGSPGZZ 583

Qy 421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPGPDAGEPGEPGADGAGPP 480
         |.|.:.||:.|| |:||:|||||..:|:|
Db 584 ---------GPPGSASGSPGKBGLBGLPGPIGPGPGPRGRTGBAGPS 631

Qy 481 GPAGPAGPPGPIG 493
         |||||| |||||
Db 632 GPPGZPGPPGPPG 644

RESULT 9

B40333
collagen alpha 1(II) chain precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
A;Accession: B40333
R;Su, M.W.; Suzuki, H.R.; Bleker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis e
A;Reference number: A40333; MUID:92011898; PMID:1918153
A;Accession: B40333
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1486 <SUA>
A;Cross-references: GB:M63595
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;37-96/Domain: von Willebrand factor type C repeat homology <VWC>

J.;Biol. Chem. 242, 2572-2574, 1967
A;Title: The incomplete hydroxylation of individual prolyl residues in collagen.
A;Reference number: A92029; MUID:67165368; PMID:4290711
A;Contents: CNBr2
A;Accession: A92029
A;Molecule type: protein
A;Residues: 20-55 <BO2>
A;Experimental source: skin and tendon
R;Butler, W.T.; Ponds, S.L.
Biochemistry 10, 2076-2081, 1971
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino a
A;Reference number: A90353; MUID:71263178; PMID:427399
A;Contents: CNBr4
A;Accession: A90353
A;Molecule type: protein
A;Residues: 56-102 <BU1>
A;Experimental source: skin
R;Butler, W.T.
Biochemistry 9, 44-50, 1970
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. The cov
A;Reference number: A90566; MUID:70085124; PMID:5411206
A;Contents: CNBr5
A;Accession: A90566
A;Molecule type: protein
A;Residues: 103-139 <BU2>
A;Experimental source: skin
R;Balian, G.; Click, E.M.; Bornstein, P.
Biochemistry 10, 4470-4478, 1971
A;Title: Structure of rat skin collagen alpha1-CB8. Amino acid sequence of the hydroxyla
A;Reference number: A90357; MUID:72136131; PMID:4335087
A;Contents: CNBr8
A;Accession: A90357
A;Molecule type: protein
A;Residues: 140-238 <BA1>
A;Experimental source: skin
R;Balian, G.; Click, E.M.; Hermodson, M.A.; Bornstein, P.
Biochemistry 11, 3798-3806, 1972
A;Title: Structure of rat skin collagen alpha1-CB8. Amino acid sequence of the hydroxyla
A;Reference number: A90362; MUID:73006942; PMID:4342027
A;Contents: CNBr8
A;Accession: A90362
A;Molecule type: protein
A;Residues: 239-418 <BA2>
A;Experimental source: skin
R;Butler, W.T.; Underwood, S.P.; Finch Jr., J.E.
Biochemistry 13, 2946-2953, 1974
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino a
A;Reference number: A90379; MUID:74271984; PMID:4366532
A;Contents: CNBr3
A;Accession: A90379
A;Molecule type: protein
A;Residues: 419-567 <BU3>
A;Experimental source: skin
R;Stoltz, M.; Timpl, R.; Furthmayr, H.; Kuehn, K.
Eur. J. Biochem. 37, 287-294, 1973
A;Title: Structural and immunogenic properties of a major antigenic determinant in neut
A;Reference number: A91209; MUID:74011954; PMID:4126850
A;Contents: CNBr6
A;Accession: A91209
A;Molecule type: protein
A;Residues: 568-651 <ST1>
A;Experimental source: skin
A;Note: this region probably corresponds to positions 949-1032 of the alpha 1(I) chain
R;Stoltz, M.; Timpl, R.; Kuehn, K.
FEBS Lett. 26, 61-65, 1972
A;Title: Non-helical regions in rat collagen alpha1-chain.
A;Reference number: A91385; MUID:73049495; PMID:4636751
A;Contents: CNBr6
A;Accession: A91385
A;Molecule type: protein
A;Residues: 651-671 <ST2>
A;Experimental source: skin

F;1258-1486/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

```
Query Match      69.9%;  Score 2147;  DB 1;  Length 1486;
Best Local Similarity  70.4%;  Pred. No. 2.3e-97;
Matches  383;  Conservative  41;  Mismatches 120;  Indels  0;  Gaps  0;

Qy     1 GSEGPEGVRGEPGPPGPAGAAGPAGPDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE   60
Db   386 GPEGAQGPRGESGTPGSPGPGASGNPGTDGIPGAKGSSGGPGIAGAPGFPGPRGPRGPGPQ  445

Qy    61 GPGGPPGKGDSGEPGAPGSKDTGAKGEPGVGVEGPGPGPAGEEGKPGARGEPGAAGPN     120
Db   446 GATGTPLGPKGETGDPGVAGFKGEQGPKGEIGSAGPQGAPGPAGEGKGRGARGEPGAAGPN  505

Qy   121 GPPGGERGAPGNRGFPGQDGLAGPKGAPGERGVPGLLGGPKGNGDPGRPGEPGLPGARGLT  180
Db   506 GPPGGERGAPGNRGFPGQDGLAGPKGAPGERGVPGLLGGPKGNGDPGRPGEPGLPGARGLT  565

Qy   181 GSPGSPGDGKTGPPGPAGEDGRPGPPGPAGEAGVMGFPGPKGAAGLPGAKGLT          240
Db   566 GRPGDAGPQGKVGPSGASGEDGRPGPPGPQGARGQPGVMGFPGPKGANGEPGKAGEKGLG   625

Qy   241 GPPGAVGPAGKDGEAGAEGPPGPAGEDGRPGPPGPAGEAGERGEEGPPGPAGPGEAGKPGEE  300
Db   626 GAPGLRGLPGKDGETGAQGPNGPAGPAGERGEQGPPGPGPSGFQGLPGPGSGEGGKPGDQ   685

Qy   301 GVPGDLGAPGPSGARGEPGPPGERGVEGPPGPAGPGADGAPGDDGAKGDAGAPGAPGSE   360
Db   686 GVPGEAGAAGLVGPRGERGFPGERGSSGPQGLQGSRGLPGTPGTDGPKGATGPSGPNGAQ  745

Qy   361 GAPGLEGMPGERGAAGLPGPKGDRGDAGPDDGAKGDAGAPGAPGSE                420
Db   746 GPPGLQGMPGERGAAGISGPKGDRGDTGEKGPEGAGKDGSRGLTGPLGPPGPAGPNGEK   805

Qy   421 GESGPSGPAGPTGARGARGDRGENGPGPGAPGFAGPPGADGQGSLKGDGQESGQKGDAGAP  480
Db   806 GESGPSGPPGPIVGARGARGDRGNGPGPAGPAGPPGADGQGSLKGDGQESGQKGDAGAP   865

Qy   481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGDATGFPGAAGRVGPPGPSGDAGPPGPGPGPA  540
Db   866 GPQGPSGAPGPQGPTGVFGPKGARGAQGPAGATGFPGAAGRVGTPGPNGPPGPGGSA   925

Qy   541 GKEG  544
Db   926 GKEG  929
```

RESULT 10
A40333
collagen alpha 1'(II) chain precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 16-Jul-1999
C;Accession: A40333
R;Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
A;Reference number: A40333; MUID:92011898; PMID:1918153
A;Accession: A40333
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1492 <SUA>
A;Cross-references: GB:M63596
A;Note: this sequence is presented as substitutions relative to another sequence in a fi
es they replace; the appropriate interpretation of the sequence figure was reconstructed
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;37-96/Domain: von Willebrand factor type C repeat homology <VWC>
F;1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

```
Query Match      69.4%;  Score 2131;  DB 2;  Length 1492;
Best Local Similarity  70.0%;  Pred. No. 1.4e-96;
Matches  381;  Conservative  41;  Mismatches 122;  Indels  0;  Gaps  0;
```

```
Qy     1 GSEGPEGVRGEPGPPGPAGAAGPAGPDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE   60
Db   389 GPEGAQGPRGESGTPGSPGPGASGNPGTDGIPGAKGSSGGPGIAGAPGFPGPRGPPGPGPQ  448

Qy    61 GPGGPPGKGDSGEPGAPGSKDTGAKGEPGVGVEGPGPGVGPAGEEGKRGARGEPGAAGPL   120
Db   449 GATGTPLGPKGETGDPGIAGFKGEHGPKGEIGSAGPQGAPGPAGEGKGRGARGEPGAAGPL  508

Qy   121 GPPGGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGPGEAGRPGEAGLPGAKGLT   180
Db   509 GPPGGERGAPGNRGFPGQDGLAGPKGAPGERGVPGLLGGPKGNGDPGRPGEPGLPGARGLT  568

Qy   181 GSPGSPGDGKTGPPGPAGEDGRPGPPGPAGEAGVMGFPGPKGAAGEAGVMGFPGPKGAAGLPGAKGLT  240
Db   569 GRPGDAGPQGKVGPSGAAGEDGRPGPPGPQGARGQPGVMGFPGPKGANGEPGKAGEKGLG  628

Qy   241 GPPGAVGPAGKDGEAGAEGPPGPAGPGADGVAGPKGAAGEGKPGEAGKPGEE          300
Db   629 GAPGLRGLPGKDGETGAQGPNGPAGPAGERGEQGPPGPGPSGFQGLPGPGSGEGGKPGDQ   688

Qy   301 GVPGDDGAPGSGARGEPGPPGERGVEGPPGPAGPGADGGDGAKGDAGAPGAPGSE          360
Db   689 GVPGEAGAAGLVGPRGERGFPGERGSSGPQGLQGSRGLPGTPGTDGPKGATGPSGPNGAQ  748

Qy   361 GAPGLEGMPGERGAAGLPGPKGDRGDAGPDDGAKGDAGAPGAPGSE                420
Db   749 GPPGLQGMPGERGAAGISGPKGDRGDTGEKGPEGAGKDGSRGLTGPLGPPGPAGPNGEK   808

Qy   421 GESGPSGPAGPTGARGAPGDRGEPGPPGARGARGSAGPPGATGFPGAAGRVGPPGPPGPGPA  480
Db   809 GESGPSGPPGPIVGARGAPGDRGENGPGPAGPAGPGSDGQAGLKGDGQESGQKGDAGAP   868

Qy   481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGDATGFPGAAGRVGPPGPSGDAGPPGPGPGPA  540
Db   869 GPQGPSGAPGPQGPTGVNGPKGARGAQGPAGATGFPGAAGRVGTPGPNGPPGPGGSA   928

Qy   541 GKEG  544
Db   929 GKEG  932
```

RESULT 11
CGBO1S
collagen alpha 1(I) chain - bovine (tentative sequence) (fragments)
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Apr-1984 #sequence_revision 31-Dec-1993 #text_change 31-Mar-2000
C;Accession: A91193; A91229; A91387; A91211; A91201; A91200; A43048; A02853
R;Rauterberg, J.; Timpl, R.; Furthmayr, H.
Eur. J. Biochem. 27, 231-237, 1972
A;Title: Structural characterization of N-terminal antigenic determinants in calf and hu
A;Reference number: A91193; MUID:72255334; PMID:4115172
A;Accession: A91193
A;Molecule type: protein
A;Residues: 1-19 <RAU>
A;Experimental source: skin
A;Note: the epsilon carbon of Lys-9, by homology with the rat alpha 1(II) chain, is conve
R;Fietzek, P.P.; Kuehn, K.
Eur. J. Biochem. 52, 77-82, 1975
A;Title: The covalent structure of collagen: amino-acid sequence of the cyanogen-bromide
A;Reference number: A91229; MUID:76022320; PMID:1164916
A;Accession: A91229
A;Molecule type: protein
A;Residues: 20-145 <FIE>
A;Experimental source: skin
A;Note: Lys-103 is hydroxylated and binds glucosylgalactose
R;Fietzek, P.P.; Wendt, P.; Kell, I.; Kuehn, K.
FEBS Lett. 26, 74-76, 1972
A;Title: The covalent structure of collagen: amino acid sequence of alpha1-CB3 from cal
A;Reference number: A91387; MUID:73049499; PMID:4673951
A;Accession: A91387
A;Molecule type: protein
A;Residues: 146-294 <FI2>
A;Experimental source: skin

R;Fietzek, P.P.; Rexrodt, F.W.; Hopper, K.E.; Kuehn, K.
Eur. J. Biochem. 38, 396-400, 1973
A;Title: The covalent structure of collagen. 2. The amino-acid sequence of alpha1-CB7 fr
A;Reference number: A91211; MUID:74086618; PMID:4359390
A;Accession: A91211
A;Molecule type: protein
A;Residues: 295-562 <FI3>
A;Experimental source: skin
R;Wendt, P.; Mark, K.V.D.; Rexrodt, F.; Kuehn, K.
Eur. J. Biochem. 30, 169-183, 1972
A;Title: The covalent structure of collagen. The amino-acid sequence of the 112 residues
A;Reference number: A91201; MUID:73042276; PMID:4343808
A;Accession: A91201
A;Molecule type: protein
A;Residues: 563-675 <WEN>
A;Experimental source: skin
R;Fietzek, P.P.; Rexrodt, F.W.; Wendt, P.; Stark, M.; Kuehn, K.
Eur. J. Biochem. 30, 163-168, 1972
A;Title: The covalent structure of collagen. Amino acid sequence of peptide alpha1-CB6-C
A;Reference number: A91200; MUID:73042275; PMID:4343807
A;Accession: A91200
A;Molecule type: protein
A;Residues: 676-758 <FI4>
A;Experimental source: skin
A;Note: Pro-726 is the only 3-hydroxyproline and the only hydroxylated proline in positi
R;Rauterberg, J.; Fietzek, P.; Rexrodt, F.; Becker, U.; Stark, M.; Kuehn, K.
FEBS Lett. 21, 75-79, 1972
A;Title: The amino acid sequence of the carboxyterminal nonhelical cross link region of
A;Reference number: A43048; MUID:...
A;Accession: A43048
A;Molecule type: protein
A;Residues: 759-779 <RA2>
A;Experimental source: skin
C;Comment: Lysines at positions 115, 124, 274, 346, 424, 496, 658, and 670 may be hydrox
C;Comment: Prolines in the third position of the tripeptide repeating unit (G-X-Y) are h
C;Comment: The order of the eight CNBr peptides in the alpha 1(I) chain of bovine skin c
9,.149, and 217 residues.
C;Comment: The complete chain contains 1052 residues.
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trimer;
F;/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match   67.6%;   Score 2076.5; DB 1; Length 779;
Best Local Similarity 78.0%;   Pred. No. 3.7e-94;
Matches 382; Conservative 18; Mismatches 51; Indels 39; Gaps 6;

Qy   55 GPSGPBPGGPGPGPKDSGEPAGPGSKGDTGAKGEPGPVGVEGPPGPAGEGKPGARGEP 114
          ||||||:|||| :||| ||| ||| ::|| :||||
Db   17 GPMGPSGPRGLPGP-------PGA-----------PGPQFQGPP---GEPGEFGASGPM 55

Qy   115 GPTGLPGPGPBERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLP 174
          |||||:|| :|||: | |||:||| || ||||||:|| :
Db   56 GPRGPPGPGKNGD--------DGEAGKPGRPGERGPGPQGAR------GLPGTAGLP 100

Qy   175 GAKGLTGSPGSPGPDGKTGPPGPAGEDGRPGSPGPAGEDGRPGPGPGPAGEPGKGAAGEPGKA 234
          | ||:|| ||||||| |:|||:||||||||||||| |:||||||||||||||||:
Db   101 GMKGHGRGFSGLDGAKDGPAGPKGEPGSPGENGAPGQMGPRGLPGFPGPGKGAAGEPGKA 160

Qy   235 GERGVPGPPGAVPGAGKDGEAGAEGPPGPAGAGERGEEGPAGSPGFEGLPGAGPPGEA 294
          |||| || |||||||| | |||||||||||||||||||: |||||||||||||
Db   161 GERGVPGPPGAVPGAGKDGEAGAQGPPGPAGERGEQGPAGSPGFQGLPGPGPPGEA 220

Qy   295 GKPGBEGVPGDLGAPGSGARGEPGFPGERGVEGPPGPAGEPGDADGAGPGDGAKGEDGAGAP 354
          ||||||||| |||:||||| | ||| |||||||||||:|||||||:|| |||||||
Db   221 GKPGEQGVPGDLGAPGDPSGARGERGFPGERGVEGPPGPAGPRGANGAPGNDGAKGDAGAP 280

Qy   355 GAPGSEGAPGLEGMPGBERGAAGLPGDGDGDGAPKGADGSPGKDGVRGLTGPIGPPGPA 414
          ||| :|||| ||||||||| ||:||:|:|:|| ||||||||
Db   281 GAPGSQGAPGLQGMPGERGAAGLPGKGDRDAGPKGADGAPGKDGVRGLTGPIGPPGPA 340

Qy   415 GAPGDKGESGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGADGEPGADGAGEPGDAGAK 474
          ||||:||| |||||| | |:|||| ||| ||||:|:|||||||:||| |:|:|| |||||
Db   341 GAPGDKGEAGPSGPA---GTRGAPGDRGEPGPPGPAGFAGPPGADGQPGAKGEPGDAGAK 397

Qy   475 GDAGPPGPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPP 534
          |||||||||||||||||||| ||||:|||| |||||||:||||||||||||||:|||||||
Db   398 GDAGPPGPAGPAGPPGPIGNVGAPGPKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPP 457

Qy   535 GPPGPAGKEG 544
          |||||||:||
Db   458 GPPGPAGKEG 467

RESULT 12
CGHU2V
collagen alpha 2 (V) chain precursor - human
C;Species: Homo sapiens (man)
C;Date: 31-Jul-1989 #sequence_revision 28-Jul-1995 #text_change 31-Dec-2000
C;Accession: A31427; A54555; S43643; A25874; I55239; I59025; A25374; A30017
R;Woodbury, D.; Benson-Chanda, V.; Ramirez, F.
J. Biol. Chem. 264, 2735-2738, 1989
A;Title: Amino-terminal propeptide of human pro-alpha2(V) collagen conforms to the stru
A;Reference number: A31427; MUID:89123368; PMID:2914927
A;Accession: A31427
A;Molecule type: mRNA
A;Residues: 1-463 <WOO>
A;Cross-references: GB:J04478; NID:g179697; PIDN:AAA51859.1; PID:g179698
A;Experimental source: placenta
R;Greenspan, D.S.; Lee, S.T.; Lee, B.S.; Hoffman, G.G.
Gene Expr. 1, 29-39, 1991
A;Title: Homology between alpha2(V) and alpha1(III) collagen promoters and evidence for
A;Reference number: A54555; MUID:92314691; PMID:1820205
A;Accession: A54555
A;Molecule type: DNA
A;Residues: 1-32 <GRE>
A;Cross-references: GB:M58529; NID:g180834; PIDN:AAC41699.1; PID:g553235
R;Moradi-Ameli, M.; Rousseau, J.C.; Kleman, J.P.; Champliaud, M.F.; Boutillon, M.M.; Be
Eur. J. Biochem. 221, 987-995, 1994
A;Title: Diversity in the processing events at the N-terminus of type-V collagen.
A;Reference number: S43642; MUID:94237164; PMID:8181482
A;Accession: S43643
A;Molecule type: protein
A;Residues: 288-291,'P',293-294,'X',296-297,606,'X',608-617 <MOR>
R;Weil, D.; Bernard, M.; Gargano, S.; Ramirez, F.
Nucleic Acids Res. 15, 181-198, 1987
A;Title: The pro alpha 2(V) collagen gene is evolutionarily related to the major fibril
A;Reference number: A25874; MUID:87146331; PMID:3029669
A;Accession: A25874
A;Molecule type: mRNA; DNA
A;Residues: 398-1496 <WEI>
A;Cross-references: GB:X04758; NID:g29588; PIDN:CAA28454.1; PID:g1340175
A;Experimental source: rhabdomyosarcoma cell line
R;Myers, J.C.; Loidl, H.R.; Stolle, C.A.; Seyer, J.M.
J. Biol. Chem. 260, 5533-5541, 1985
A;Title: Partial covalent structure of the human alpha 2 type V collagen chain.
A;Reference number: I55239; MUID:85182703; PMID:2985598
A;Accession: I55239
A;Molecule type: mRNA
A;Status: translated from GB/EMBL/DDBJ
A;Residues: 1002-1226 <RE2>
A;Cross-references: GB:M10956; NID:g180427; PIDN:AAA52007.1; PID:g180428
A;Note: part of this sequence were determined by protein sequencing
R;Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long ar
A;Reference number: I59025; MUID:85216505; PMID:3858826
A;Accession: I59025
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1003-1034 <RES>
A;Cross-references: GB:M11135; NID:g179693; PIDN:AAA51857.1; PID:g179694
A;Note: part of this sequence were determined by protein sequencing
R;Myers, J.C.; Loidl, H.R.; Seyer, J.M.; Dion, A.S.
J. Biol. Chem. 260, 11216-11222, 1985
A;Title: Complete primary structure of the human alpha-2 type V procollagen COOH-termin
A;Reference number: A25374; MUID:85289337; PMID:2411731

A;Accession: A25374
A;Molecule type: mRNA
A;Residues: 1227-1417,'T',1419-1437,'S',1439-1496 <MYE>
A;Cross-references: GB:M11718; NID:g180912; PIDN:AAA52058.1; PID:g180913
A;Experimental source: normal fibroblasts
R;Tsipouras, P.; Schwartz, R.C.; Liddell, A.C.; Salkeld, C.S.; Weil, D.; Ramirez, F.
Genomics 3, 275-277, 1988
A;Title: Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2, located on
A;Reference number: A30017; MUID:89138450; PMID:3224983
A;Accession: A30017
A;Molecule type: DNA
A;Gene: GDB:COL5A2
A;Residues: 1449-1463,'E',1465-1495,'A' <TSI>
A;Cross-references: GB:J03051; NID:g179695; PIDN:AAA51858.1; PID:g179696
A;Note: the authors translated the codon GAA for residue 1460 as Gln, and GAG for residu
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
are 5-hydroxylated and subsequently O-glycosylated.
C;Comment: The amino-terminal propeptide domain appears not to be completely cleaved.
C;Genetics:
A;Gene: GDB:COL5A2
A;Cross-references: GDB:119064; OMIM:120190
A;Map position: 2q31-2q31
A;Introns: 33/1; 812/3; 830/3; 848/3; 884/3; 902/3; 922/3; 974/3; 1046/3; 1064/3; 1448/3
A;Complex: type V collagen may be a homotrimer of alpha 1(V) chains (see PIR:CGHU1V), a
alpha 2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among the
length, is formed with desmosine cross-links made from lysine and allysine residues
C;Function:
A;Description: structural component of extracellular fibrous polymer associated with cel
A;Note: may play a role in controlling the lateral growth of collagen I fibrils
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproli
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-193/Product: collagen alpha 2(V) chain #status predicted <MAT>
F;27-193/Domain: amino-terminal propeptide (uncleaved) #status predicted <NPP>
F;27-108/Region: nonhelical
F;40-99/Domain: von Willebrand factor type C repeat homology <VWC>
F;109-186/Region: helical
F;187-208/Region: nonhelical
F;209-1225/Region: helical
F;505-505/Region: cell attachment (R-G-D) motif
F;941-943/Region: cell attachment (R-G-D) motif
F;1064-1066/Region: cell attachment (R-G-D) motif
F;1067-1069/Region: cell attachment (R-G-D) motif
F;1097-1099/Region: cell attachment (R-G-D) motif
F;1124-1126/Region: cell attachment (R-G-D) motif
F;1133-1135/Region: cell attachment (R-G-D) motif
F;1225-1250/Region: carboxy-terminal nonhelical telopeptide
F;1251-1496/Domain: carboxyl-terminal propeptide #status predicted <CPP>
F;1269-1496/Domain: fibrillar collagen carboxyl-terminal homology<FCC>
F;27/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F;193-194/Cleavage site: Ala-Gln (procollagen N-endopeptidase) #status predicted
F;194/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicte
F;201/Modified site: allysine (Lys) #status predicted
F;290,299,296,608,614,1004,1007,1013,1028,1034/Modified site: 4-hydroxyproline (Pro) #st
F;299,1139/Modified site: 5-hydroxylysine (Lys) #status predicted
F;299,1139/Binding site: carbohydrate (Lys) (covalent) #status predicted
F;1025/Modified site: 5-hydroxylysine (Lys) #status experimental
F;1250-1251/Cleavage site: Glu-Asp (procollagen C-endopeptidase) #status predicted
F;1259,1397/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1293,1299,1325/Disulfide bonds: interchain #status predicted
F;1333-1494,1402-1447/Disulfide bonds: intrachain #status predicted

Query Match   62.6%; Score 1921.5; DB 1; Length 1496;
Best Local Similarity 64.0%; Pred. No. 1.9e-86;
Matches 348; Conservative 40; Mismatches 153; Indels 3; Gaps 1;

```
Qy     1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGADGAPGIAGAPGFPGARGPSGPE 60
           ||| ||| |::||||   ||  :|||||  ||        ||  ||  || ||: |
Db   396 GPEGPGQRGETGPPGPVGSPGLPGAIGTDGTTPGRGPTSPGT---SGP--PGSACPPGSP 452

Qy    61 GPGGPPGPKGDSGGEPGAPGSKGDTGAKGEPGPGVGVEGPPGPAGEBGKPGARGEPGPTGLP 120
           ||| ||::|:   ||| :| ||  :||  ||  ||  ::|| ||:  |::|:|||||| ||
Db   453 GPQGSTGPQGNSGLPGDPGFKGEAGPKGEPGPHGIQGPIGGPPGEBGKGKRGPRGDPGTLGPP 512
```

```
Qy   181 GSPGSPGDDGKTGPPGPAGEDGRPGPPGPPGPGARGEAGVMGFPGPKGAAGEPGKAGERGVP 240
           | ||:   ||:||| | || ||  |||| || || |||| ||||| |: |: ||| || ||
Db   573 GNPGVGQPEGKLGPLGAPGEDGRPGPPGSIGIKGQPGTMGLPGPKGSNGDPGKPGEAGNP 632

Qy   241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGEREGEEGPAGSPGPEGLPGAPGPGEAGKPGEE 300
           | || :| ||:|| :| :|| |  | |  ::|::| || ||  | ||  | |:|||||
Db   633 GVPGGQRGAPGKDGKVGPYGPPGPPGLRGERGEQGPGPTGPFQGHPGPPGPEGGKPGDQ 692

Qy   301 GVPGDLGAPGPSGARGBPGFPGERGVEGPPGPGAPGPDGGAKGDAGAPGAPGSE 360
           |||| ||||| ||| ||  || | :||| || | ||  |||||| ||||| ||
Db   693 GVPGGPGAVGPLGPRGERGNPGERGEPGITGLPGEKGMAGGHGPDGPKGSPGPSGTPGDT 752

Qy   361 GAPGLEMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
           | || :||| | :|||| ||: || :| ||  |:|  ||  ||  | || ||| ||
Db   753 GPPGLQGMPGERGIAGTPGPKGDRGIGEKGAEGTAGNDGAGGLPGPLGPPGPAGLLGEK 812

Qy   421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPGEPGDGEPGAKGBPGDAGAKGDAGPP 480
           | |||| || | ||| || ||:| |  ||| | |  :| |:| :||||||||  ||||:|
Db   813 GEPGPRGLVGPPGSRGNPGSRGBNGTGAVGPAGPQGSDGQPGVKGBPGEPGQKGDAGSP 872

Qy   481 GPAGPAGPPGPIGDVGPAGAKGARGSAGPPGATGPGAAGRVGPPGPGSGDAGPPGPGPA 540
           || |:|| || | | || :| ||||||| |:| ||| |||||||  || |||| |  |
Db   873 GPQGLAGSPGPHGPNGVPGLKGGRGTGPPGATGFPGSAGRVGPPGPAGAPGPAGPIGEP 932

Qy   541 GKEG 544
           ||||
Db   933 GKEG 936
```

RESULT 13
A43291
collagen alpha 2(I) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A43291; A54328
R;Phillips, C.L.; Morgan, A.L.; Lever, L.W.; Wenstrup, R.J.
Genomics 13, 1345-1346, 1992
A;Title: Sequence analysis of a full-length cDNA for the murine pro alpha 2(I) collagen
A;Reference number: A43291; MUID:92372043; PMID:1505972
A;Accession: A43291
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1373 <PHI>
A;Cross-references: GB:X58251; NID:g50488; PIDN:CAA41205.1; PID:g50489
A;Note: sequence extracted from NCBI backbone (NCBIP:112027)
R;Phillips, C.L.; Lever, L.W.; Pinnell, S.R.; Quarles, L.D.; Wenstrup, R.J.
J. Invest. Dermatol. 97, 980-984, 1991
A;Title: Construction of a full-length murine Proalpha2(I) collagen cDNA by the polymer
A;Reference number: A54328; MUID:92084969; PMID:1748823
A;Accession: A54328
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-110 <PH2>
C;Genetics:
A;Gene: COL1A2
C;Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;1145-1373/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match   62.4%; Score 1915; DB 1; Length 1373;
Best Local Similarity 64.2%; Pred. No. 3.7e-86;
Matches 349; Conservative 45; Mismatches 150; Indels 0; Gaps 0;

```
Qy     1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE 60
           |:|| |||:|| ||  ||:  ||  |
Db   280 GPAGPAGPRGEVGLPGLSGPVGPPGNPGTNGLTGAKGATGLPGVAGADPLPGPAEGEPGPGPA 339

Qy    61 GPGGPPGPKGDSGGEPGAPGSKGDTGAKGEPGPGVGVEGPPGPAGEBGKPGARGEPGPTGLP 120
           ||| ||::| |:|: |: | |   |:  | | ||  ::||  || :||| |: ||||||
Db    61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPGVGVEGPPGPAGEBGKPGARGEPGPTGLP 120
```

```
Db   340  GAAGATGARGLVGPGPGPAGSKGESGNKGEPGSVGAQGPPGPSGEGKRGSPGEAGSAGPA  399
                                                                                    
Qy        |:||   |||:||||||   ||||||||:||:|:|||:|||||||||||||||||||||||
Db   574  GNPGVQGPEGKLGPLGAPGEDGRPGPPGSIGRGQPGSMGVPGPKGSSGDLGKPGEAGNA  633

Qy   121  GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSSPGAGPKGSPGEAGLPGAKGLT  180
                                                                                         
Db        |||                                                              
Qy   574  GNPGVQGPEGKLGPLGAPGEDGRPGPPGSIGRGQPGSMGVPGPKGSSGDLGKPGEAGNA  

Db   400  GPPGLRGSPGSRGLPGADGRAGVWGPPGNRGSTGPAGIRGPNGDAGRPGEPGLMGPRGLP  459
                                                                                    
Qy        |||::||::|:|::||||||::|:||:|||||||||||||||   |||||:|||::
Qy   241  GPPGAVGPAGKDGEAGABGPPGPAGPAGEREGEGPAGSPGFEGLPGPAGPGEAGKPGEE  300

Qy   181  GSPGSPGPDGKTGPPGPAGEDGRPGPPGPGERGEGAGVWGFPGPKGAAGEPGKAGERGVP  240
                                                                                    
Db        ||||: ||                                                         
Qy   634  GVPGQRGAPGKDGEVGPSGPVGPSPGVPPLLAGERGEAGPPGPGPTGPQGLPGPGSGKAGDQ  693

Db   460  GSPGNVGPSGGKEGPVGLPGIDGRFPGPIGPAGPRGEAGNIGFPGPKGPGSGDPGKPGERGHP  519
                                                                                    
Qy        |||: ||||  |                                                     
Qy   301  GVPGDLGAPGPSGARGEPGFPGPGERGVEGPPGPGPADGAPGDDGAKGDAGAPGAPGSE  360

Db        |||: ||                                                          
Db   694  GVPGEPGAVGPLGPRGRGNPGERGEPGITGLPGEKGMAGGHGPDGPKGNPGPTGTIGDT  753

Qy   241  GPPGAVGPAGKDGEAGABGPPGPAGPAGEREGEGPAGSPGFEGLPGPAGPGEAGKFGEE  300
                                                                                    
Db        |||                                                              
Qy   361  GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGDGVRGLTGPGPPGPAGAPGDK  420

Db   520  GLAGARGAPGPDGNNGAQGPGPGPQGVQGGKGEQGPAGPPGFQGLLGPSGTTGEVGKPGER  579
                                                                                    
Db   754  GPPGLQGMPGERGIAGTPGPKGDRGGIGEKGAEGTAGNDGARGLPGPLGPPGPAGLLGAP  813

Qy   301  GVPGDLGAPGSGARGEPGPGERGVEGPPGPGPADGAPGDDGAKGDAGAPGAPGSE  360
                                                                                    
Db        |:|||:|||||:||||:||:||::|:||:|||||:||||||||||:|||||:
Qy   421  GESGPSGPAGPTGARGEPGEPGPGPAGRGEPGPPGPAGDGEPGAKGDAGAKGDAGPP  480

Db   580  GLPGEFGLPGAPGPRGERGTPGESGAAGPSGPIGSRGPSGAPGDGNKGEAGAVGAPGSA  639
                                                                                    
Db   814  GEPGPRGLVGPPGGRGNPGSRGENGPTGAVGPAGFGPQGSDGQPGVKGEPGQQGDAGSP  873

Qy   361  GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPGPPGPAGAPGDK  420
                                                                                    
Db        |:||: ::|||:|||                                                  
Qy   481  GPAGPAGPPGPIGDVGAPGAKGARGSAGPGDGATGFPGAAGRVGPPGPSGDAGPPGPGPGA  540

Db   640  GASGPGLPGERGAAGIPGGKGEKGETGLRGDTGNTGRDGARGIPGAVGAPGPAGASGDR  699
                                                                                    
Db   874  GPQGLAGSPGPHGPHGVPGLKGGRGTQGPPGATGFPGSAGRVGPPGPAGAPGPAGEP  933

Qy   421  GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGDGEPGAKGEPGDAGAKGDAGPP  480
                                                                                    
Db        ||                                                               
Qy   700  GEAGAAGPSGPGAPRGSPGERGEVGPAGPNGFAGPAGAAGQPGAKGEKGTKGPKGENGIV  759

Qy   481  GPAGPAGPPGPIGDVGAPGAKGARGSAGPGDGGPP2MTGPFGAAGRTGPPGPVGPSGIAGPPGPGPPGAA  819
                                                                                    
Qy   760  GPTGSVGAAGPSGPNGPGPNGPPGPVGSRGDGGPP2MTGPFGAAGRTGPPGPVGPSGIAGPPGPGPPGAA

Qy   541  GKEG  544
          ||||
Db   820  GKEG  823

Qy   541  GKEG  544
          ||||
Db   934  GKEG  937
```

RESULT 14
I49607
procollagen type V alpha 2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
C;Accession: I49607
R;Andrikopoulos, K.; Suzuki, H.R.; Solursh, M.; Ramirez, F.
Dev. Dyn. 195, 113-120, 1992
A;Title: Localization of pro-alpha 2(V) collagen transcripts in the tissues of the devel
A;Reference number: I49607; MUID:93214071; PMID:1297453
A;Accession: I49607
A;Molecule type: mRNA
A;Residues: 1-1497 <RES>
A;Cross-references: GB;L02918; NID:g309180; PIDN:AAA37440.1; PID:g309181
C;Genetics:
A;Gene: Col5a-2
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
F;39-98/Domain: von Willebrand factor type C repeat homology <VWC>
F;1270-1497/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match   62.2%; Score 1910; DB 2; Length 1497;
Best Local Similarity   63.8%; Pred. No. 6.9e-86;
Matches 347; Conservative 39; Mismatches 158; Indels 0; Gaps 0;

```
Qy    1  GSGPGEVGREPGPPGPAGPAGDGEPGAKGDAGAPGAPGFPGARGPSGPE  60
                                                                
Db  394  GEBGPQQGRGETGPPGPAGSQGLPGAVGTDGTPGRKGATGSAGTSGPPGLAGPPGGSPGPQ  453

Qy   61  GPPGPPGKGDSGEPGAPGSKGDTGAKGEPGVGVEGPPGPAGEEGKGPGARGEPGPGPTGLP  120
                                                                         
Db  454  GSTGPQGIRGQGSDGPGVPGFKGAGPPGKGEPGPHGIGQGPIGPPGEEGKCRGRPGPGDPG7VGPP  513

Qy  121  GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT  180
                                                                         
Db  514  GMGERGAPGNRGFPGSDGLPGPKGAQGERGPVGSSGPKGQGQGDPGRPGEPGLPGARGLT  573

Qy  181  GSPGSPGPDGKTGPPGPAGEDGRPGPPGPAGEDGRPGPPGGARGEAGVMGFPGPKGAAGEPGKAGERGVP  240
```

```
Qy      1 GSEGPEGVRGEPGPPGPAGAGPAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE   60
            ||||            |||     :| |   | ||||     ||||||   ||
Db    359 GSNGSPGQRGEPGPQGHAGAQGPPGPPGNNGSPGGKGEMGPAGIPGAPGLIGARGPPGPA  418

Qy     61 GPGGPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEEGKPGARGEPGPTGLP  120
            ||||    | :  :||  | :: ||:||| | |    ||    |::  : :||
Db    419 GTNGIPGTRGPSGEPGKNGAKGEPGARGERGEAGSPGIPGPKGBEDGKDGSPGEPGANGLP  478

Qy    121 GPPGERGCPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT  180
            ||||    |  :: :::||     ||||      ||    | ||||  :  :::
Db    479 GAAGERGPSGFRGPAGPNGIPGEKGPPGERGGPGPAGPRGVAGEPGRDGTPGGPGIRGMP  538

Qy    181 GSPGSPGPDGKTGPPGEPAGEDGRPGPPGPPGPPGARGEAGVMGFPGPKGAAGEPGKAGERGVP  240
            ||||    | |||| | |||||||   ||  :: :||:  :  |||||||||||||
Db    539 GSPGGPGNDGKPGPPGSQGESGRPGPPGPSGPPGPGPSGPRGQPGVMGFPGPKGNDGAPGKNGERGGP  598

Qy    241 GPPGAVGPAGKDGEAGAEGPPGAPGPAGERGEEGPAGSPGFEGLPGPAGPPGEAGKPGEB  300
            ||||  | ||||  ||    | :||:  :: :||:|||   |||||||  ||||||
Db    599 GGPGLPGPAGKNGETGPQGPPGPTGPACDKGDSGPPGPGPGQLQGIPGTCGPPGENGKPGEP  658

Qy    301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGADGA---PGDDGAKGDAGAPGAP  357
            | :: :   :| :||| | | ||  | ||| | :: ::  |::    ||  | :
Db    659 GPKGEVGAPGAPGGKGDSGAPGER---GPPGTAGIPGARGGAGDGVPGKDGPRGPAGPQP  715

Qy    358 GSEGAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGHKDGVRGLTGPIGPPGPAGAP  417
            |:::||||     |  |:|:| |  :|:  :|||   | |||  | |||    |||
Db    716 GASGSPGLQGMPGERGGPGSPGKGEKGEPGGAGADGVPGKDGPRGPAGPIGPPGPAGQP  775

Qy    418 GDKGESGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGPGADGEPGAKGEPGDAGAKGDA  477
            | |: |||   || ||   | || |||| || |  :|| ||||| || |||:
Db    776 GDKGEGGSSPGGLPGIAGPRGGPGERGEHGPPGAGFPGPAGPGQNGEPGAKGERGAPGEKGEG  835

Qy    478 GPPGPARGPAGPPGPIGDVGAGPAGKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPPGPP  537
            ||||  |            ||   | |    | |||||     ||| |  :  || |
Db    836 GPPGPAGPTGSSGPAGPPGPQGPGQVKGERGSPGPPGPTAGFPGGRGLPGPPGNNGNPGPPGPS  895

Qy    538 GPAGKEG  544
            |   ||:|
Db    896 GAPGKDG  902
```

Search completed: May 7, 2004, 16:30:21
Job time : 22 secs

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    May 7, 2004, 16:24:19 ; Search time 18 Seconds
                                     (without alignments)
                              1573.676 Million cell updates/sec

Title:    US-10-658-989A-4
Perfect score: 3070
Sequence: 1 GSEGPEGVRGEPGPPGPAGA.........PGPSGDAGPPGPPGPAGKEG 544

Scoring table:  BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:   141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:   141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                          SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2966	96.6	1464	1	CA11_HUMAN	P02452 homo sapien
2	2940	95.8	1460	1	CA11_CANFA	Q9xsj7 canis famil
3	2857	93.1	1453	1	CA11_MOUSE	P11087 mus musculu
4	2775	90.4	1453	1	CA11_CHICK	P02457 gallus gall
5	2202	71.7	1418	1	CA12_HUMAN	P02458 homo sapien
6	2192	71.4	1459	1	CA12_MOUSE	P28481 mus musculu
7	2151	70.1	671	1	CA11_RAT	P02454 rattus norv
8	2076.5	67.6	779	1	CA11_BOVIN	P02453 bos taurus
9	1949	63.5	1364	1	CA21_BOVIN	P02465 bos taurus
10	1934.5	63.0	1262	1	CA13_CHICK	P12105 gallus gall
11	1934	63.0	1372	1	CA21_RAT	P02466 rattus norv
12	1921.5	62.6	1496	1	CA25_HUMAN	P05997 homo sapien
13	1915	62.4	1372	1	CA21_MOUSE	Q01149 mus musculu
14	1905.5	62.1	1049	1	CA13_BOVIN	P04258 bos taurus
15	1901	61.9	1466	1	CA13_HUMAN	P02461 homo sapien
16	1899	61.9	1464	1	CA12_HUMAN	P08121 homo sapien
17	1898	61.8	1366	1	CA21_CANFA	O46592 canis famil
18	1880	61.2	1366	1	CA21_HUMAN	P08123 homo sapien
19	1834.5	59.8	1355	1	CA21_RANCA	042350 rana catesb
20	1822	59.3	1362	1	CA21_CHICK	P02467 gallus gall
21	1789	58.3	1356	1	CA21_ONCMY	093484 oncorhynchu
22	1496	48.7	747	1	CA12_BOVIN	P02459 bos taurus
23	1454.5	47.4	1838	1	CA15_HUMAN	P20908 homo sapien
24	1449	47.2	1806	1	CA1B_HUMAN	P12107 homo sapien
25	1447	47.1	1745	1	CA35_HUMAN	P25940 homo sapien
26	1432	46.6	1804	1	CA1B_MOUSE	Q61245 mus musculu
27	1423.5	46.4	1736	1	CA2B_HUMAN	P13942 homo sapien
28	1414.5	46.1	1736	1	CA2B_MOUSE	Q64739 mus musculu
29	1410	45.9	1027	1	CAFF_RIFPPA	P30754 riftia pach
30	1387	45.2	1685	1	CA54_HUMAN	P29400 homo sapien
31	1375.5	44.8	2944	1	CA17_HUMAN	Q02388 homo sapien
32	1372	44.7	1763	1	CA24_ASCSU	P27393 ascaris suu
33	1357.5	44.2	1690	1	CA44_HUMAN	P53420 homo sapien
34	1356	44.2	911	1	CA1B_BOVIN	Q28083 bos taurus
35	1344.5	43.8	1670	1	CA34_HUMAN	Q01955 homo sapien
36	1339.5	43.6	1603	1	CA1F_HUMAN	Q07092 homo sapien
37	1320	43.0	1669	1	CA14_HUMAN	P02462 homo sapien
38	1310.5	42.7	1758	1	CA24_CABEL	P17140 caenorhabdi
39	1310	42.7	1669	1	CA14_MOUSE	P02463 mus musculu
40	1286.5	41.9	1712	1	CA24_HUMAN	P08572 homo sapien
41	1281.5	41.7	1516	1	CA1H_HUMAN	P39060 homo sapien
42	1279.5	41.7	754	1	CA54_CANFA	Q28247 canis famil
43	1277	41.6	684	1	CA39_HUMAN	Q14050 homo sapien
44	1276.5	41.6	1143	1	CA1I_HUMAN	Q14993 homo sapien
45	1276	41.6	921	1	CA19_MOUSE	Q05722 mus musculu

```
                          ALIGNMENTS

RESULT 1
CA11_HUMAN
ID   CA11_HUMAN     STANDARD;      PRT;   1464 AA.
AC   P02452; Q14037; Q15176;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Collagen alpha 1(I) chain precursor.
GN   COL1A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-472 FROM N.A.
RX   MEDLINE=89025644; PubMed=177843;
RA   Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA   Jaenisch R., Prockop D.J.;
RT   "Structure of a full-length cDNA clone for the prepro alpha 1(I)
RT   chain of human type I procollagen.";
RL   Biochem. J. 253:919-922(1988).
RN   [2]
RP   SEQUENCE OF 1-181 FROM N.A.
RX   MEDLINE=84270697; PubMed=6462220;
RA   Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
RA   Myers J., Williams C., Ramirez F.;
RT   "Human pro alpha 1(I) collagen gene structure reveals evolutionary
RT   conservation of a pattern of introns and exons.";
RL   Nature 310:337-340(1984).
RN   [3]
RP   SEQUENCE OF 162-301.
RC   TISSUE=Skin;
RX   MEDLINE=71038625; PubMed=5529814;
RA   Click E.M., Bornstein P.;
RT   "Isolation and characterization of the cyanogen bromide peptides from
RT   the alpha 1 and alpha 2 chains of human skin collagen.";
RL   Biochemistry 9:4699-4706(1970).
RN   [4]
RP   SEQUENCE OF 263-268.
RC   TISSUE=Skin;
RX   MEDLINE=71001508; PubMed=4319110;
RA   Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
RT   "A comparative study of glycopeptides derived from selected
RT   vertebrate collagens. A possible role of the carbohydrate in fibril
RT   formation.";
RL   J. Biol. Chem. 245:5042-5048(1970).
RN   [5]
RP   SEQUENCE OF 425-1464 FROM N.A.
RX   MEDLINE=84080385; PubMed=6689127;
RA   Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
RA   Prockop D.J.;
RT   "Nucleotide sequences of complementary deoxyribonucleic acids for the
RT   pro alpha 1 chain of human type I procollagen. Statistical evaluation
RT   of structures that are conserved during evolution.";
RL   Biochemistry 22:5213-5223(1983).
RN   [6]
```

RP   SEQUENCE OF 1229-1454 FROM N.A.
RC   TISSUE=Bone;
RX   MEDLINE=81244208; PubMed=3340531;
RA   Maekelae J.K., Raassina M., Virta A., Vuorio E.;
RT   "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
RT   domain.";
RL   Nucleic Acids Res. 16:349-349(1988).
RN   [7]
RP   SEQUENCE OF 1-34 FROM N.A.
RX   MEDLINE=88097389; PubMed=3480516;
RA   Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
RT   "Regulatory elements in the first intron contribute to
RT   transcriptional control of the human alpha 1(I) collagen gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
RN   [8]
RP   SEQUENCE OF 1-34 FROM N.A.
RX   MEDLINE=81130970; PubMed=2857713;
RA   Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.;
RT   "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.";
RL   J. Biol. Chem. 260:2315-2320(1985).
RN   [9]
RP   SEQUENCE OF 1-44 FROM N.A.
RX   MEDLINE=88033098; PubMed=2822714;
RA   Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
RA   de Wet W.J.;
RT   "DNA sequences in the first intron of the human pro-alpha 1(I)
RT   collagen gene enhance transcription.";
RL   J. Biol. Chem. 262:15151-15157(1987).
RN   [10]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=91184577; PubMed=2010058;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in collagen genes: causes of rare and some common diseases
RT   in humans.";
RL   FASEB J. 5:2052-2060(1991).
RN   [11]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97255959; PubMed=9101290;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT   associated collagen (type IX) and network-forming collagen (type X)
RT   cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL   Hum. Mutat. 9:300-315(1997).
RN   [12]
RP   REVIEW ON OI VARIANTS.
RX   MEDLINE=91374476; PubMed=1895312;
RA   Byers P.H., Wallis G.A., Willing M.C.;
RT   "Osteogenesis imperfecta: translation of mutation to phenotype.";
RL   J. Med. Genet. 28:433-442(1991).
RN   [13]
RP   REVIEW ON OI VARIANTS.
RX   MEDLINE=97169389; PubMed=9016532;
RA   Dalgleish R.;
RT   "The human type I collagen mutation database.";
RL   Nucleic Acids Res. 25:181-187(1997).
RN   [14]
RP   VARIANT OI-II CYS-1166.
RX   MEDLINE=86287390; PubMed=3016737;
RA   Cohn D.H., Byers P.H., Steinmann B., Gelinas R.E.;
RT   "Lethal osteogenesis imperfecta resulting from a single nucleotide
RT   change in one human pro alpha 1(I) collagen allele.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).
RN   [15]
RP   VARIANT OI-II ARG-569.
RX   MEDLINE=87222295; PubMed=3108247;
RA   Bateman J.F., Chan D., Walkers I.D., Rogers J.G., Cole W.G.;
RT   "Lethal perinatal osteogenesis imperfecta due to the substitution of
RT   arginine for glycine at residue 391 of the alpha 1(I) chain of type I
RT   collagen.";
RL   J. Biol. Chem. 262:7021-7027(1987).
RN   [16]
RP   VARIANT OI-II CYS-926.

RX   MEDLINE=88013031; PubMed=3667599;
RA   Vogel B.E., Minor R.R., Freund M., Prockop D.J.;
RT   "A point mutation in a type I procollagen gene converts glycine 748
RT   of the alpha 1 chain to cysteine and destabilizes the triple helix in
RT   a lethal variant of osteogenesis imperfecta.";
RL   J. Biol. Chem. 262:14737-14744(1987).
RN   [17]
RP   VARIANT OI-II ARG-842.
RX   MEDLINE=88298828; PubMed=3403550;
RA   Bateman J.F., Lamande S.R., Dahl H.H., Chan D., Cole W.G.;
RT   "Substitution of arginine for glycine 664 in the collagen alpha 1(I)
RT   chain in lethal perinatal osteogenesis imperfecta. Demonstration of
RT   the peptide defect by in vitro expression of the mutant cDNA.";
RL   J. Biol. Chem. 263:11627-11630(1988).
RN   [18]
RP   VARIANT OI CYS-1195.
RX   MEDLINE=89218628; PubMed=3244312;
RA   Labhard M.E., Wirtz M.K., Pope F.M., Nicholls A.C., Hollister D.W.;
RT   "A cysteine for glycine substitution at position 1017 in an alpha
RT   1(I) chain of type I collagen in a patient with mild dominantly
RT   inherited osteogenesis imperfecta.";
RL   Mol. Biol. Med. 5:197-207(1988).
RN   [19]
RP   VARIANT OI-II VAL-434.
RX   MEDLINE=89254493; PubMed=2470760;
RA   Patterson E., Smiley E., Bonadio J.;
RT   "RNA sequence analysis of a perinatal lethal osteogenesis imperfecta
RT   mutation.";
RL   J. Biol. Chem. 264:10083-10087(1989).
RN   [20]
RP   VARIANT OI-IV SER-1010.
RX   MEDLINE=89308591; PubMed=2745420;
RA   Marini J.C., Grange D.K., Gottesman G.S., Lewis M.B., Koeplin D.A.;
RT   "Osteogenesis imperfecta type IV. Detection of a point mutation in
RT   one alpha 1(I) collagen allele (COL1A1) by RNA/RNA hybrid analysis.";
RL   J. Biol. Chem. 264:11893-11900(1989).
RN   [21]
RP   VARIANTS OI-II ALA-1106; VAL-1151; ARG-1154 AND VAL-1184.
RX   MEDLINE=89380165; PubMed=2777764;
RA   Lamande S.R., Dahl H.-H.M., Cole W.G., Bateman J.F.;
RT   "Characterization of point mutations in the collagen COL1A1 and
RT   COL1A2 genes causing lethal perinatal osteogenesis imperfecta.";
RL   J. Biol. Chem. 264:15809-15812(1989).
RN   [22]
RP   VARIANT OI SER-1022.
RX   MEDLINE=90062068; PubMed=2511192;
RA   Pack M., Constantinou C.D., Kalia K., Nielsen K.B., Prockop D.J.;
RT   "Substitution of serine for alpha 1(I)-glycine 844 in a severe
RT   variant of osteogenesis imperfecta minimally destabilizes the triple
RT   helix of type I procollagen. The effects of glycine substitutions on
RT   thermal stability are either position of amino acid specific.";
RL   J. Biol. Chem. 264:19694-19699(1989).
RN   [23]
RP   VARIANT OI-II CYS-1082.
RX   MEDLINE=89109573; PubMed=2913053;
RA   Constantinou C.D., Nielsen K.B., Prockop D.J.;
RT   "A lethal variant of osteogenesis imperfecta has a single base
RT   mutation that substitutes cysteine for glycine 904 of the alpha 1(I)
RT   chain of type I procollagen. The asymptomatic mother has an
RT   unidentified mutation producing an overmodified and unstable type I
RT   procollagen.";
RL   J. Clin. Invest. 83:574-584(1989).
RN   [24]
RP   VARIANTS OI-II CYS-272; CYS-704 AND CYS-896.
RX   MEDLINE=90009313; PubMed=2794057;
RA   Starman B.J., Eyre D., Charbonneau H., Harrylock M., Weis M.A.,
RA   Weiss L., Graham J.M., Byers P.H.;
RT   "Osteogenesis imperfecta. The position of substitution for glycine by
RT   cysteine in the triple helical domain of the pro alpha 1(I) chains of
RT   type I collagen determines the clinical phenotype.";
RL   J. Clin. Invest. 84:1206-1214(1989).
RN   [25]
RP   VARIANT OI-II CYS-422.

```
Query Match            96.6%;   Score 2966;  DB 1;  Length 1464;
Best Local Similarity  95.6%;   Pred. No. 2.5e-113;
Matches  520;  Conservative  21;  Mismatches   3;  Indels   0;  Gaps   0;

Qy      1 GSEGPEGVRGEPGPPGPAGAAGPGPAGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE 60
            |||::||| || ||||||||:||||||||||||||||| ||||| ||:|||||||||||
Db    362 GSEGPQGVRGEPGPPGPAGAAGPAGADGAPGAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQ 421

Qy     61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEEGKPGARGEPGPTGLP 120
            |||||||||||:|||||||||||||||||||| |||||||||::|||||||||||||||
Db    418 GPSGPPGPKGNSGEPGAPGNKGDTGAKGEPGPTGIQGPPGAGEEGKRGARGEPGPTGLP 477

Qy    121 GPGGERGGPGSRGFPGADGVAGPKGPAGERGSGPGAPGKGSPGEAGRPGEAGLPGAKGLT 180
            |||||||||:|:|||||||||||||||||||||||||||||||||||||||||||||||
Db    478 GPGGERGGPGSRGFPGADGVAGPKGPAGERGSGPGAPGKGSPGEAGRPGEAGLPGAKGLT 537

Qy    181 GSPGSPGDGKTGPPGPAGEDGRPGPPGPAGERGEAGVMGFPGPKGAAGEPGKAGERGVP 240
            ||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db    538 GSPGSPGDGKTGPPGPAGQDGRPGPPGPAGERGQAGVMGFPGPKGAAGEPGKAGERGVP 597

Qy    241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEEGPAGSPGFEGLPGPAGPAGAPGDK 300
            |||||||||||||||:|||||||||||||:|||||||||||:||:|||||||||||
Db    598 GPPGAVGPAGKDGEAGAQGPPGPAGSPGFQGLPGPAGPAGEQGEQ 657

Qy    301 GVPGDLGAPGSGARGEPGFPGERGVEGPPGPAGERGEEGPAGSPGKDGVRGLTGPIGPPGPAGAPGDK 360
            ||||||||:|:|||||||||:|||:|||||||||:|||||||||||||||||||||:||||||||||
Db    658 GVPGDLGAPGSGARGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQ 717

Qy    361 GAPGLEGMPGERGAAGLPGPKGDRDGAPGKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
            |||||||||||||||:||||||||||||||||||||||||||||||||||||||||||:
Db    718 GAPGLQGMPGERGAAGLPGPKGDRDGAPGKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 777
```

```
DR   EMBL; U08020; AAA88912.1; -.
DR   EMBL; X15896; CAA33904.1; -.
DR   EMBL; M14423; AAA37333.1; -.
DR   EMBL; M17491; AAA37334.1; -.
DR   EMBL; X06753; CAA29927.1; -.
DR   EMBL; K03036; AAA37332.1; JOINED.
DR   EMBL; K03029; AAA37332.1; JOINED.
DR   EMBL; K03030; AAA37332.1; JOINED.
DR   EMBL; K03031; AAA37332.1; JOINED.
DR   EMBL; K03032; AAA37332.1; JOINED.
DR   EMBL; K03033; AAA37332.1; JOINED.
DR   EMBL; K03034; AAA37332.1; JOINED.
DR   EMBL; K03035; AAA37332.1; JOINED.
DR   PIR; S57243; S21626.
DR   MGD; MGI:88467; Col1a1.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR002181; Fibrinogen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD000007; Clg_helix; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal.
FT   SIGNAL        1     22
FT   PROPEP       23    151       AMINO-TERMINAL PROPEPTIDE.
FT   CHAIN       152   1207       COLLAGEN ALPHA 1(I) CHAIN.
FT   PROPEP     1208   1453       CARBOXYL-TERMINAL PROPEPTIDE.
FT   DOMAIN       29     87       VWFC.
FT   DOMAIN      152    167       NONHELICAL REGION (N-TERMINAL).
FT   DOMAIN      168   1181       TRIPLE-HELICAL REGION.
FT   DOMAIN     1182   1207       NONHELICAL REGION (C-TERMINAL).
FT   CARBOHYD     56     56       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1354   1354       N-LINKED (GLCNAC...) (POTENTIAL).
FT   SITE        734    736       CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE       1082   1084       CELL ATTACHMENT SITE (POTENTIAL).
FT   CONFLICT   1450   1450       A -> V (IN REF. 5).
SQ   SEQUENCE   1453 AA;  137944 MW;  3B802E535DF81808 CRC64;

Query Match            93.1%;  Score 2857;  DB 1;  Length 1453;
Best Local Similarity  91.4%;  Pred. No. 6e-109;
Matches  497;  Conservative  28;  Mismatches  19;  Indels  0;  Gaps  0;

Qy      1  GSEGPEGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPGPE  60
           ||:|||  |||||||||||:|:||||:||||||||||||||||:||||||||||||::|||
Db    351  GSEGPQGVRGEPGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPGPQ  410

Qy     61  GPGPGPGPKGDSGPGPAGPSKGDTGAKGEPGPVCVEGPPGPAGEBEGKPGARGEPGPTGLP  120
           ||||:|||||:||||||||:|||:|||||||||:||||||||||||:||||||||||||||
Db    411  GPSGPGPKGNSGEPGAPGNKGDTGAKGEPGATGVQGPPGPAGEBGKRGARGEPGPSGLP  470

Qy    121  GPPGERGGPGSRGPFPGADGVAGPKGPAGBRGSPGPAGRGSPGPAGPKGSPGEAGLPGAKGLT  180
           ||||:||||||||||||||||||||||||:||||:|||||||||||:||||||||||||||
Db    471  GPPGERGGPGSRGFPGADGVAGPKGPAGPKGSPGEAGRPGEAGLPGAKGLT  530

Qy    181  GSPGSPGPDGKTGPPGPAGEDGREPGPPGPAGVMGFPGPKGAAGEPGKAGERGVP  240
           ||||:|||||||||||||||||||||:||||:||||||||||||||||||||:
Db    531  GSPGSPGPDGKTGPPGPAGQDGRPGAGPPGPAGVMGFPGPKGTAGEPGTAGERGLP  590
```

```
Qy    421  GESGPSGPAGPTGARGAPGDRGREPGPPGPAGFAGPPGADGEPGAKGEPGDAGAKGDAGPP  480
           ||:||||||:||||||||||||:|||:||||||:|||||:|||||||:|||||:|||||||
Db    778  GEAGPSGPAGPTGARGAPGDRGEBEPGPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPP  837

Qy    481  GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPPGPPGPA  540
           ||:|||||||||::|||:||||||||||||:||||||||||||||:|||||||||||||||
Db    838  GPAGPTGPPGPIGNVGAPGKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPA  897
```

```
Qy    541  GKEG  544
           ||||
Db    898  GKEG  901
```

```
RESULT 3
CA11_MOUSE
ID   CA11_MOUSE     STANDARD;     PRT;  1453 AA.
AC   P11087; Q60635;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Collagen alpha 1(I) chain precursor.
GN   COL1A1 OR COLA1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N;
RX   MEDLINE=96033240; PubMed=8535610;
RA   French B.T., Lee W.-H., Maul G.G.;
RT   "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)
RT   collagen protein.";
RL   Gene 39:311-312(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96033240; PubMed=8535610;
RA   Li S.W., Khillan J., Prockop D.J.;
RT   "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
RT   of type I procollagen.";
RL   Matrix Biol. 14:593-595(1995).
RN   [3]
RP   SEQUENCE OF 735-1130 FROM N.A.
RX   MEDLINE=83141374; PubMed=6298597;
RA   Monson J.M., Friedman J., McCarthy B.J.;
RT   "DNA sequence analysis of a mouse pro alpha 1 (I) procollagen gene:
RT   evidence for a mouse B1 element within the gene.";
RL   Mol. Cell. Biol. 2:1362-1371(1982).
RN   [4]
RP   SEQUENCE OF 518-1128 FROM N.A.
RX   MEDLINE=86137403; PubMed=3841523;
RA   Monson J.M., McCarthy B.J.;
RT   "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
RT   evidence for insertions or deletions in gene coding sequences.";
RL   DNA 1:59-69(1981).
RN   [5]
RP   SEQUENCE OF 1442-1453 FROM N.A.
RX   MEDLINE=88124276; PubMed=3340560;
RA   Mooslehner K., Harbers K.;
RT   "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size
RT   of the 3'-untranslated region.";
RL   Nucleic Acids Res. 16:773-773(1988).
CC   -!- FUNCTION: Type I collagen is a member of group I collagen
CC       (fibrillar forming collagen).
CC   -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC   -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC       bones. In bones the fibrils are mineralized with calcium
CC       hydroxyapatite.
CC   -!- PTM: Prolines at the third position of the tripeptide repeating
CC       unit (G-X-Y) are hydroxylated in some or all of the chains.
CC   -!- SIMILARITY: Contains 1 VWFC domain.
```

Qy 241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEBGPAGSPGFEGLPGPAGGPPGEAGKPGEE 300
Db 591 GPPGAVGPAGKDGEAGAQGAPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQ 650

Qy 301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGADGAPGDDGAKDGAPGAPGSE 360
Db 651 GVPGDLGAPGPSGARGERGFPGERGVQGPPGPAGPRGNNGAPGNDGAKDGTGAPGAPGSQ 710

Qy 361 GAPGLEGMPGERGAAGLPGPKGDRDGDAGPKGADGSPGKDGVRGLTGPIGPGPAGAPGDK 420
Db 711 GAPGLQGMFGERGAAGLPGPKGDRDGDAGPKGADGSPGKDGARGLTGPIGPGPPAGAPGDK 770

Qy 421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGAGEPGAKGEPGDAGAKGDAGPP 480
Db 771 GEAGPSGPPGPTGARGAPGDRGEAGPPGPAGDGQPGAKGEPGDTGVKGDAGPP 830

Qy 481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPGSGDAGPPGPPGPA 540
Db 831 GPAGPAGPPGPIGNVGAPGPKGPRGAAGPPGATGFPGAAGRVGPPGPGSGNAGPPGPGPPGPV 890

Qy 541 GKEG 544
Db 891 GKEG 894

RESULT 4
ID   CA1_CHICK   STANDARD;   PRT;   1453 AA.
AC   P02457;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Collagen alpha 1(I) chain precursor.
GN   COL1A1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE OF 1-153 FROM N.A.
RX   MEDLINE=88056316; PubMed=3678834;
RA   Finer M.H., Boedtker H., Doty P.;
RT   "Construction and characterization of cDNA clones encoding the 5' end
RT   of the chicken pro alpha 1(I) collagen mRNA.";
RL   Gene 56:71-78(1987).
RN   [2]
RP   SEQUENCE OF 1-144 FROM N.A.
RX   MEDLINE=88007542; PubMed=2820966;
RA   Finer M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RT   "Unusual DNA sequences located within the promoter region and the
RT   first intron of the chicken pro-alpha 1(I) collagen gene.";
RL   J. Biol. Chem. 262:13323-13332(1987).
RN   [3]
RP   SEQUENCE OF 152-1187.
RX   MEDLINE=82231995; PubMed=7093229;
RA   Highberger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M.,
RA   Kang A.H., Gross J.;
RT   "Amino acid sequence of chick skin collagen alpha 1(I)-CB8 and the
RT   complete primary structure of the helical portion of the chick skin
RT   collagen alpha 1(I) chain.";
RL   Biochemistry 21:2048-2055(1982).
RN   [4]
RP   SEQUENCE OF 1200-1205.
RX   MEDLINE=72243016; PubMed=5047697;
RA   Eyre D.R., Glimcher M.J.;
RT   "Evidence for a previously undetected sequence at the carboxyterminus
RT   of the alpha 1 chain of chicken bone collagen.";
RL   Biochem. Biophys. Res. Commun. 48:720-726(1972).
RN   [5]
RP   SEQUENCE OF 981-1453 FROM N.A.
RX   MEDLINE=81160715; PubMed=6927845;
RA   Fuller F., Boedtker H.;
RT   "Sequence determination and analysis of the 3' region of chicken pro-
RT   alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
RT   including the carboxy-terminal propeptide sequences.";
RL   Biochemistry 20:996-1006(1981).
RN   [6]
RP   SEQUENCE OF 1311-1453 FROM N.A.
RX   MEDLINE=80134546; PubMed=6987088;
RA   Showalter A.M., Pesciotta D.M., Eikenberry E.F., Yamamoto T.,
RA   Pastan I., Decrombrugghe B., Fietzek P.P., Olsen B.R.;
RT   "Nucleotide sequence of a collagen cDNA-fragment coding for the
RT   carboxyl end of pro alpha 1(I)-chains.";
RL   FEBS Lett. 111:61-65(1980).
CC   -!- FUNCTION: Type I collagen is a member of group I collagen
CC   (fibrillar forming collagen).
CC   -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC   -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC   bones. In bones the fibrils are mineralized with calcium
CC   hydroxyapatite.
CC   -!- PTM: Proline residues at the third position of the tripeptide
CC   repeating unit (G-X-Y) are hydroxylated in some or all of the
CC   chains. Pro-1153 is the only 3-hydroxypro and the only
CC   hydroxylated proline in position X.
CC   -!- SIMILARITY: Contains 1 VWFC domain.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; M17839; AAA48704.1; -.
DR   EMBL; M17838; AAA48704.1; JOINED.
DR   EMBL; V00401; CAA23695.1; -.
DR   EMBL; M10571; AAA48671.1; ALT_SEQ.
DR   EMBL; M17607; AAA48672.1; -.
DR   PIR; A27179; A27179.
DR   PIR; I50629; I50629.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR002181; Fibrinogen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD000007; Clg_helix; 2.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal; Pyrrolidone carboxylic acid.
FT   SIGNAL   1   22
FT   PROPEP   23   151   AMINO-TERMINAL PROPEPTIDE.
FT   CHAIN   152   1205   COLLAGEN ALPHA 1(I) CHAIN.
FT   PROPEP   1206   1453   C-TERMINAL PROPEPTIDE.
FT   DOMAIN   31   89   VWFC.
FT   MOD_RES   152   152   PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES   254   254   HYDROXYLATION (POTENTIAL).
FT   MOD_RES   851   851   HYDROXYLATION (POTENTIAL).
FT   MOD_RES 1081   1081   HYDROXYLATION (POTENTIAL).
FT   MOD_RES 1097   1097   HYDROXYLATION (POTENTIAL).
FT   MOD_RES 1153   1153   HYDROXYLATION.
FT   MOD_RES 1187   1187   HYDROXYLATION.
FT   CONFLICT 1187   1187   F -> L (IN REF. 5).
FT   CONFLICT 1441   1441   Q -> H (IN REF. 6).
SQ   SEQUENCE   1453 AA; 137789 MW; 3BC615213421F4D CRC64;

Query Match   90.4%;   Score 2775; DB 1; Length 1453;
Best Local Similarity 89.0%;   Pred. No. 1.2e-105;
Matches 484; Conservative 28; Mismatches 32; Indels 0; Gaps 0;

```
Qy    1  GSEGPEGVRGPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGPGARGPSGPE   60
Db  351  GSEGPQSRGPGPPGPAGANPGADGQPGAKGATGAPGIAGAPGPGARGPSGPQ        410

Qy   61  GPGGPPGPGKGDSGEPGAPGSKDTGAKGEPGPVGPGPPGPAGEBGKPGARGEPGTGLP  120
Db  411  GPSGAPGPKGNSGEPGAPGNKGDTGAKGEPGPAGVCQGPPGPAGEBGKRGARGEPGAGLP 470

Qy  121  GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
Db  471  GPAGERGAPGSRGFPGADGIAGPKGPPGERGSPGEAGRPGEAGLPGAKGLT          530

Qy  181  GSPGSPGDGKTGPGPPGAGEDGRPGPPGPPGPAGEGRGVMGFPGPKGAAGEPGKAGERGVP 240
Db  531  GSPGSPGDGKTGPGPPGAGQDGRPGPAGPGARGQAGVMGFPGPKGAAGEPGKPGERGAP  590

Qy  241  GPPGAVGPAGKDGEAGABGPPGPAGPAGERGEBGPAGSPGFEGLPGPAGPPGEAGKPGEE 300
Db  591  GPPGAVGAAGKDGEAGAQGPPGPGTGPAGERGEBGQGPAGPGFGQLPGPAGPPGEAGKPGEQ 650

Qy  301  GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGADGAPGDDGAKGDAGAPGPGSE  360
Db  651  GVPGNAGAPGPAGARGERGFPGERGVQGPPGPQGPRGANGAPGNDGAKGDAGAPGAPGNE 710

Qy  361  GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 420
Db  711  GPPGLEGMPGERGAAGLPGAKGDRGDPGPKGADGAPGKDGLRGLTGPIGPPGPAGAPGDK 770

Qy  421  GESGPSGPAGPTGARGAPGDRGEBGPPGPPGAFGAFGPGADGEPGAKGEBPGDAGAKGDAGPP 480
Db  771  GEAGPPGPAGPTGARGAPGDRGEPGPPGPPGAFGAFGPGADGQPGAKGETGDAGAKGDAGPP 830

Qy  481  GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGSDAGPPGPGPA    540
Db  831  GPAGPTGAPGPAGZVGAPGPKGARGSAGPPGATGFPGAAGRVGPPGSGNIGLPGPPGPA  890

Qy  541  GKEG 544
Db  891  GKZG 894
```

RESULT 5
CA12_HUMAN
ID   CA12_HUMAN   STANDARD;   PRT;   1418 AA.
AC   P02458;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Collagen alpha 1(II) chain precursor [Contains: Chondrocalcin].
GN   COL2A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87031574; PubMed=3021582;
RA   Nunez A.M., Kohno K., Martin G.R., Yamada Y.;
RT   "Promoter region of the human pro-alpha 1(II)-collagen gene.";
RL   Gene 44:11-16(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90067946; PubMed=2587267;
RA   Su M.W., Lee B., Ramirez F., Machado M., Horton W.;
RT   "Nucleotide sequence of the full length cDNA encoding for human type
RT   II procollagen.";
RL   Nucleic Acids Res. 17:9473-9473(1989).
RN   [3]
RP   SEQUENCE OF 1-28 FROM N.A.
RN   [4]
RP   SEQUENCE OF 432-1145 FROM N.A.
RA   Ramirez F.;
RL   Submitted (DEC-1988) to the EMBL/GenBank/DDBJ databases.

RP   SEQUENCE OF 963-1418 FROM N.A.
RX   MEDLINE=85190534; PubMed=3857598;
RA   Cheah K.S.E., Stoker N.G., Griffin J.R., Grosveld F.G., Solomon E.;
RT   "Identification and characterization of the human type II collagen
RT   gene (COL2A1).";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:2555-2559(1985).
RN   [5]
RP   SEQUENCE OF 1120-1398 FROM N.A.
RX   MEDLINE=85306861; PubMed=3840017;
RA   Elima K., Maekelae J.K., Vuorio T., Kauppinen S., Knowles J.,
RA   Vuorio E.;
RT   "Construction and identification of a cDNA clone for human type II
RT   procollagen mRNA.";
RL   Biochem. J. 229:183-188(1985).
RN   [6]
RP   SEQUENCE OF 1106-1418 FROM N.A.
RX   MEDLINE=88067771; PubMed=2825137;
RA   Elima K., Vuorio T., Vuorio E.;
RT   "Determination of the single polyadenylation site of the human pro
RT   alpha 1(II) collagen gene.";
RL   Nucleic Acids Res. 15:9499-9504(1987).
RN   [7]
RP   SEQUENCE OF 1227-1289 FROM N.A.
RX   MEDLINE=88104139; PubMed=3002437;
RA   Nunez A.M., Francomano C., Young M.F., Martin G.R., Yamada Y.;
RT   "Isolation and partial characterization of genomic clones coding for
RT   a human pro-alpha 1 (II) collagen chain and demonstration of
RT   restriction fragment length polymorphism at the 3' end of the gene.";
RL   Biochemistry 24:6343-6348(1985).
RN   [8]
RP   SEQUENCE OF 1176-1226 FROM N.A.
RX   MEDLINE=84118798; PubMed=6320112;
RA   Strom C.M., Upholt W.B.;
RT   "Isolation and characterization of genomic clones corresponding to
RT   the human type II procollagen gene.";
RL   Nucleic Acids Res. 12:1025-1038(1984).
RN   [9]
RP   SEQUENCE OF 35-167 FROM N.A.
RX   MEDLINE=89233118; PubMed=2714801;
RA   Su M.W., Benson-Chanda V., Vissing H., Ramirez F.;
RT   "Organization of the exons coding for pro alpha 1(II) collagen N-
RT   propeptide confirms a distinct evolutionary history of this domain of
RT   the fibrillar collagen genes.";
RL   Genomics 4:438-441(1989).
RN   [10]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=91184577; PubMed=2010058;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in collagen genes: causes of rare and some common diseases
RT   in humans.";
RL   FASEB J. 5:2052-2060(1991).
RN   [11]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97255955; PubMed=9101290;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT   associated collagen (type IX), and network-forming collagen (type X)
RT   cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL   Hum. Mutat. 9:300-315(1997).
RN   [12]
RP   VARIANT SER-1074.
RX   MEDLINE=90036909; PubMed=2572591;
RA   Vissing H., D'Alessio M., Lee B., Ramirez F., Godfrey M.,
RA   Hollister D.W.;
RT   "Glycine to serine substitution in the triple helical domain of pro-
RT   alpha 1 (II) collagen results in a lethal perinatal form of short-
RT   limbed dwarfism.";
RL   J. Biol. Chem. 264:18265-18267(1989).
RN   [13]
RP   VARIANT SEDC 1095-GLY--TYR-1330 DEL.
RX   MEDLINE=89266907; PubMed=2543071;
RA   Lee B., Vissing H., Ramirez F., Rogers D., Rimoin D.;
RT   "Identification of the molecular defect in a family with

RT spondyloepiphyseal dysplasia.";
RL Science 244:978-980(1989).
RN [14]
RP VARIANT OSTEOARTHRITIS CYS-650.
RX MEDLINE=90370826; PubMed=1975693;
RA Ala-Kokko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;
RT "Single base mutation in the type II procollagen gene (COL2A1) as a
RT cause of primary osteoarthritis associated with a mild
RT chondrodysplasia.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568(1990).
RN [15]
RP VARIANT OI-IV VAL-717.
RX MEDLINE=91291136; PubMed=2064612;
RA Bateman J.F., Hannagan M., Chan D., Cole W.G.;
RT "Characterization of a type I collagen alpha 2(I) glycine-586 to
RT valine substitution in osteogenesis imperfecta type IV. Detection of
RT the mutation and prenatal diagnosis by a chemical cleavage method.";
RL Biochem. J. 276:765-770(1991).
RN [16]
RP VARIANT OSTEOARTHRITIS CYS-650.
RX MEDLINE=91086471; PubMed=1985108;
RA Eyre D.R., Weis M.A., Moskowitz R.W.;
RT "Cartilage expression of a type II collagen mutation in an inherited
RT form of osteoarthritis associated with a mild chondrodysplasia.";
RL J. Clin. Invest. 87:357-361(1991).
RN [17]
RP VARIANT HYPOCHONDROGENESIS GLU-984.
RX MEDLINE=93054548; PubMed=1429602;
RA Bogaert R., Tiller G.E., Wies M.A., Gruber H.E., Rimoin D.L.,
RA Cohn D.H., Eyre D.R.;
RT "An amino acid substitution (Gly853-->Glu) in the collagen alpha
RT 1(II) chain produces hypochondrogenesis.";
RL J. Biol. Chem. 267:22522-22526(1992).
RN [18]
RP VARIANT HYPOCHONDROGENESIS SER-705.
RX MEDLINE=92262484; PubMed=1374906;
RA Horton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,
RA Ramirez F., Vitale E., Lee B.;
RT "Characterization of a type II collagen gene (COL2A1) mutation
RT identified in cultured chondrocytes from human hypochondrogenesis.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).
RN [19]
RP VARIANT WS-II ASP-198.
RX MEDLINE=93304428; PubMed=8317498;
RA Koerkko J., Ritvaniemi P., Haataja L., Kaeaeriaeinen H.,
RA Kivirikko K.I., Prockop D.J., Ala-Kokko L.;
RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate
RT for glycine alpha 1-67 and that causes cataracts and retinal
RT detachment: evidence for molecular heterogeneity in the Wagner
RT syndrome and the Stickler syndrome (arthro-ophthalmopathy).";
RL Am. J. Hum. Genet. 53:55-61(1993).
RN [20]
RP VARIANT SEMD CYS-840.
RA Tiller G.E., Weis M.A., Lachman R.S., Cohn D.H., Rimoin D.L.,
RA Eyre D.R.;
RT "A dominant mutation in the type II collagen gene (COL2A1) produces
RT spondyloepimetaphyseal dysplasia (SEMD), Strudwick type.";
RL Am. J. Hum. Genet. 53:A209-A209(1993).
RN [21]
RP VARIANT OSTEOARTHRITIS CYS-650.
RX MEDLINE=93282819; PubMed=8807190;
RA Holderbaum D., Malemud C.J., Moskowitz R.W., Haqqi T.M.;
RT "Human cartilage from late stage familial osteoarthritis transcribes
RT type II collagen mRNA encoding a cysteine in position 519.";
RL Biochem. Biophys. Res. Commun. 192:1169-1174(1993).
RN [22]
RP VARIANT SEMD ARG-285.
RX MEDLINE=93252400; PubMed=8486375;
RA Vikkula M., Ritvaniemi P., Vuorio A.F., Kaitila I., Ala-Kokko L.,
RA Peltonen L.;
RT "A mutation in the amino-terminal end of the triple helix of type II
RT collagen causing severe osteochondrodysplasia.";
RL Genomics 16:282-285(1993).

RN [23]
RP VARIANT SEDC CYS-206.
RX MEDLINE=94063862; PubMed=8244341;
RA Williams C.J., Considine E.L., Knowlton R.G., Reginato A., Neumann G.,
RA Harrison D., Buxton P., Jimenez S.A., Prockop D.J.;
RT "Spondyloepiphyseal dysplasia and precocious osteoarthritis in a
RT family with an Arg75-->Cys mutation in the procollagen type II gene
RT (COL2A1).";
RL Hum. Genet. 92:499-505(1993).
RN [24]
RP VARIANT SEDC CYS-920.
RX MEDLINE=93315508; PubMed=8325895;
RA Chan D., Taylor T.K.F., Cole W.G.;
RT "Characterization of an arginine 789 to cysteine substitution in
RT alpha 1(II) collagen chains of a patient with spondyloepiphyseal
RT dysplasia.";
RL J. Biol. Chem. 268:15238-15245(1993).
RN [25]
RP VARIANT SEDC SER-1128.
RX MEDLINE=93140139; PubMed=8423604;
RA Cole W.G., Hall R.K., Rogers J.G.;
RT "The clinical features of spondyloepiphyseal dysplasia congenita
RT resulting from the substitution of glycine 997 by serine in the alpha
RT 1(II) chain of type II collagen.";
RL J. Med. Genet. 30:27-35(1993).

```
Query Match      71.7%;  Score 2202;  DB 1;  Length 1418;
Best Local Similarity   71.9%;  Pred. No. 1.3e-82;
Matches   391;  Conservative   40;  Mismatches  113;  Indels    0;  Gaps    0;

Qy    1 GSEGPEGVRGEPGPPGPAGAAGPAGPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE    60
          ::  | ||||||||| |:  ||   ||  ||| ||||  |||||||||||||||:  ||
Db  315 GPEGAQGPGRGEPGTPGSPGPAGASGNPGTDGIPGAKGSAGAPGIAGAPGFPGPRGRGPPDPQ  374

Qy   61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVBGPGPGPAGEBGKPGARGEPGPTGLP   120
          | ||  |||| |||  | | ||  ||||||||| |||  | |||||| |||  |:  |:
Db  375 GATGPLPGKGQTGKPGIAGPKGEQGPKGEPGPAGERGSPGEAGRPGSGEAGRPGAGERGVGPI   434

Qy  121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGAKGLT   180
          |||| | |  | ||  |||  | |||  ||| ||| ||| | | |||||| | |:  |||||
Db  435 GPPGERGAPGNRGFPGQDGLAGPKGAPGERGPSGLAGPKGANGDPGRPGEPGLPGARGLT   494

Qy  181 GSPGSGPGDGKTGPFGPAGEDGRPGPGPPGAGAAGEPGKAGERGVP   240
          | ||  |  || | ||  || | ||  |||   |  ||| |||  |
Db  495 GRPGDAGPGQGKVGPSGAPGEDGRPGPPGPQGPGQCARGQGPVMGFPGPKGANGEBGKPGAGEKGLP   554

Qy  241 GPPGAVGPAGKDGEAGAEGPPGPAGPGAGERGEEGPAGSPGFEGLPGAGPPGGEAGKPGEE   300
          ||||  || |||   |   || |||  |||||  | |  ||  || || |   || ||  ||
Db  555 GAPGLRGLPGKDGETGAEGGPPGPGPAGERGEBQGAPGSGGFQGGLPGPPGPBGGGKPGDQ   614

Qy  301 GVPGDIGAPGPSGARGEPGFPGERGVEGPPGPPGPAGPPGPADGAGPGDDGAKGDAGAPGSE   360
          |||| ||| ||| ||| | |||| | |  |  ||||| || |  || ||| || | | |
Db  615 GVPGEAGAPGLVGPRGERGPPGERGSPGAQLQGPRGLPGTPGTDGPKGASGPAGPPGAQ   674

Qy  361 GAPGLEGMPGERGAAGLPGPKGDRGDAGAPGKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK   420
          |  || | ||||   |    || | | | | |  |  || | |:|||| ||||| ||| |:
Db  675 GPPGLQGMPGERGAAGIAGPKGDRGDVGEKGPEGAPGKDGRGLTGPIGPPGPAGANGEK   734

Qy  421 GESGPSGPAGPTGARGAPGDRGEPGPPGADGEPGAKGBPGDAGAKGDAGPP   480
          | || || ||| ||| ||| |||||| | || || |||  || || ||  |
Db  735 GEVGPPGPAGSAGARGAPGERGETGPPGTSGIAGPPGADGQPGAKGBQGARGQKDGAGAP   794

Qy  481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPPGPPGPA   540
          | ||   | |  |  |  || ||| |  ||  |||  |:  | |||| | || | | | ||
Db  795 GPQGPSGAPGPQGPTGPVTGPKGARGAQGPPGATGFPGAAGRVGPPGSNGNPGPPGPPGPS   854

Qy  541 GKEG   544
          || :|
Db  855 GKDG   858
```

RESULT 6
CA12_MOUSE

```
ID   CA42_MOUSE     STANDARD;      PRT;   1459 AA.
AC   P28481;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Collagen alpha 1(II) chain precursor [Contains: Chondrocalcin].
GN   COL2A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=91358489; PubMed=1885613;
RA   Metsaranta M., Toman D., de Crombrugghe B., Vuorio E.;
RT   "Mouse type II collagen gene. Complete nucleotide sequence, exon
RT   structure, and alternative splicing.";
RL   J. Biol. Chem. 266:16862-16869(1991).
RN   [2]
RP   SEQUENCE OF 1455-1459 FROM N.A.
RX   MEDLINE=91274355; PubMed=2054384;
RA   Metsaranta M., Toman D., de Crombrugghe B., Vuorio E.;
RT   "Specific hybridization probes for mouse type I, II, III and IX
RT   collagen mRNAs.";
RL   Biochim. Biophys. Acta 1089:241-243(1991).
CC   -!- FUNCTION: Collagen type II is specific for cartilaginous tissues.
CC   -!- SUBUNIT: Homotrimers of alpha 1(II) chains.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Long;
CC         IsoId=P28481-1; Sequence=Displayed;
CC       Name=Short;
CC         IsoId=P28481-2; Sequence=VSP_001139, VSP_001140;
CC   -!- PTM: Prolines at the third position of the tripeptide repeating
CC       unit (G-X-Y) are hydroxylated in some or all of the chains.
CC   -!- SIMILARITY: Contains 1 VWFC domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M65161; AAA68100.1; -.
DR   EMBL; X57982; CAA41047.1; -.
DR   MGD; MGI:88452; Col2a1.
DR   InterPro; IPR008161; Clg.helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000885; Fib-collagen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD000007; Clg_helix; 5.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal; Alternative splicing.
FT   SIGNAL        1     25       POTENTIAL.
FT   PROPEP       26    153       AMINO-TERMINAL PROPEPTIDE
FT                                (BY SIMILARITY).
FT   CHAIN       154   1213       COLLAGEN ALPHA 1(II) CHAIN.
FT   PROPEP     1214   1459       CHONDROCALCIN.
FT   DOMAIN       32     89       VWFC.
FT   DOMAIN      173   1186       TRIPLE-HELICAL REGION.
FT   DOMAIN     1187   1213       NONHELICAL REGION (C-TERMINAL).
FT   VARSPLIC     29     29       Q -> R (in isoform Short).
FT                                /FTId=VSP_001139.
```

```
FT   VARSPLIC     30     98       Missing (in isoform Short).
FT                                /FTId=VSP_001140.
SQ   SEQUENCE   1459 AA;  139154 MW;  F6C84FA7C532E7F2 CRC64;

Query Match             71.4%;  Score 2192;  DB 1;  Length 1459;
Best Local Similarity   71.7%;  Pred. No. 3.4e-82;
Matches  390;  Conservative  40;  Mismatches 114;  Indels  0;  Gaps  0;

Qy         1 GSGPGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE   60
             ||| : ||||||||| || ||| |:|: :||  | |||| | :: :|||||||| |
Db       356 GPEGAQGSRGEPGNPGSSPGPAGASGNPGTDGIPGAKGSAGAPGIAGAPGFPGPRGPGPGPQ  415

Qy        61 GPGGPPGKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGAGEEGKPGAGEPGPTGLP  120
             ||| ||||| ||||||||||| |:  |||  ||||  |||||||||||| ||||
Db       416 GATGPLGPKGQAGEPGIAGFKGDQGPKGETGPAGPQGAPGPAGEGKRGARGEPGGAGPI  475

Qy       121 GPPGERGPSGRGFPGADGVAGPKGPAGERGSPGPAGERSSPGPAGGPGEAGLPGAKGLT  180
             ||| || |  |  || |||   ||||| ||||  |  ||||  | :|||||||
Db       476 GPPGERGAPGNGFPGGQDGLAGPKGAPGERGPSGLAGPKGANGDPGRPGEGLPGARGLT  535

Qy       181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGARGEAGVMGPPGPKGAAGEPGKAGERGVP  240
             |||| ||||| |||||||  || |||||||||| |   |||||||||| |||| || ::
Db       536 GRPGDAGPGQGKVGPSGAPGEDGRPGPPGPPGQQGARGQPGVMGPFGPKGANGEPGAGEKGLA  595

Qy       241 GPPGAVGPAGKDGEAGAEGPPGPAGEGREEGPAGSPGEGLPGPAGPPGAGKPGEE  300
             || ||  |||| ||||  || || || |  || ||   ||  ||| ||||||
Db       596 GAPGLRGLPGKDGETGAAGPPGSPGAGEROQGAPGSPGQGLPGPPGPGEGGQQGDQ  655

Qy       301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPPGADGAPGDDGAKGDAGAPGAPGSE  360
             |: |  ||  ||   | ||||| |  | ||| ||  || ||  ||  | || | ||
Db       656 GPEEAGAPGLVGPRGERGFPGERGSPGAQGLQGPRGLPGVTGVTGTDGPKGAAGPDGPPGAQ  715

Qy       361 GGPGLEGMPGERGAAGLPGPKGDRGAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  420
             | | |  |||| ||| || || ||  ||    |||| ||:  | ||||||| |:
Db       716 GPGLQGMPGERGAAGIAGPKGDRGDVGEKGPEGAPGKDGGRGLTGPIGPPGPAGANGEK  775

Qy       421 GESGPSGPAGPTGARGARGPPGPAGEPGPPGPAGFAGPEGPDGAKGDAGPP  480
             | |||||| ||  |||  || |:|   |: :|||||| :
Db       776 GEAGPPGSGSTGARGAPGEPGETGPPGPAGFAGPPGPAGRVGPPGPSGDAGPPGPGPA  835

Qy       481 GPAGPAGPPGPIGDVGAPGAKGARGSAGPPGPATGPPGAKGRVGPPGPPGGPPGPA  540
             | |||:|| || |  |  :|||| |  || :  ||||| |: : ||||| |
Db       836 GPQPSGAPGPQQGPTGVTGPKGARGAQGPPGPGATGFPGAAGRVGPPGANGPNPGAGPPGPA  895

Qy       541 GKEG  544
             |:||
Db       896 GKDG  899
```

```
RESULT 7
CA11_RAT
ID   CA11_RAT      STANDARD;      PRT;    671 AA.
AC   P02454; P02455;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Collagen alpha 1(I) chain (Fragments).
GN   COL1A1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-19.
RX   MEDLINE=69155173; PubMed=5777344;
RA   Bornstein P.;
RT   "Comparative sequence studies of rat skin and tendon collagen. II.
RT   The absence of a short sequence at the amino terminus of the skin
RT   alpha-1 chain.";
RL   Biochemistry 8:63-71(1969).
RN   [2]
RP   SEQUENCE OF 5-19.
RX   MEDLINE=67162268; PubMed=5337886;
```

```
RA   Kang A.H., Bornstein P., Piez K.A.;
RT   "The amino acid sequence of peptides from the cross-linking region of
RT   rat skin collagen.";
RL   Biochemistry 6:788-795(1967).
RN   [3]
RP   SEQUENCE OF 20-55.
RX   MEDLINE=67165368; PubMed=4290711;
RA   Bornstein P.;
RT   "The incomplete hydroxylation of individual prolyl residues in
RT   collagen.";
RL   J. Biol. Chem. 242:2572-2574(1967).
RN   [4]
RP   SEQUENCE OF 56-102.
RX   MEDLINE=71263178; PubMed=4327399;
RA   Butler W.T., Ponds S.L.;
RT   "Chemical studies on the cyanogen bromide peptides of rat skin
RT   collagen. Amino acid sequence of alpha 1-CB4.";
RL   Biochemistry 10:2076-2081(1971).
RN   [5]
RP   SEQUENCE OF 103-139.
RX   MEDLINE=70085124; PubMed=5411206;
RA   Butler W.T.;
RT   "Chemical studies on the cyanogen bromide peptides of rat skin
RT   collagen. The covalent structure of alpha 1-CB5, the major
RT   hexose-containing cyanogen bromide peptide of alpha 1.";
RL   Biochemistry 9:44-50(1970).
RN   [6]
RP   SEQUENCE OF 140-238.
RX   MEDLINE=72136131; PubMed=4335087;
RA   Balian G., Click E.M., Bornstein P.;
RT   "Structure of rat skin collagen alpha 1-CB8. Amino acid sequence of
RT   the hydroxylamine-produced fragment HA1.";
RL   Biochemistry 10:4470-4478(1971).
RN   [7]
RP   SEQUENCE OF 239-418.
RX   MEDLINE=73006942; PubMed=4342027;
RA   Balian G., Click E.M., Hermodson M.A., Bornstein P.;
RT   "Structure of rat skin collagen alpha 1-CB8. Amino acid sequence of
RT   the hydroxyl amine-produced fragment HA2.";
RL   Biochemistry 11:3798-3806(1972).
RN   [8]
RP   SEQUENCE OF 419-567.
RX   MEDLINE=74271984; PubMed=4366532;
RA   Butler W.T., Underwood S.P., Finch J.E. Jr.;
RT   "Chemical studies on the cyanogen bromide peptides of rat skin
RT   collagen. Amino acid sequence of alpha 1-CB3.";
RL   Biochemistry 13:2946-2953(1974).
RN   [9]
RP   SEQUENCE OF 568-651.
RX   MEDLINE=74011954; PubMed=4126850;
RA   Stoltz M., Timpl R., Furthmayr H., Kuehn K.;
RT   "Structural and immunogenic properties of a major antigenic
RT   determinant in neutral salt-extracted rat-skin collagen.";
RL   Eur. J. Biochem. 37:287-294(1973).
RN   [10]
RP   SEQUENCE OF 651-671.
RX   MEDLINE=73049495; PubMed=4636751;
RA   Stoltz M., Timpl R., Kuehn K.;
RT   "Non-helical regions in rat collagen alpha 1-chain.";
RL   FEBS Lett. 26:61-65(1972).
RN   [11]
RP   SEQUENCE OF 529-567 FROM N.A.
RX   MEDLINE=85122694; PubMed=6395893;
RA   Genovese C., Rowe D., Kream B.;
RT   "Construction of DNA sequences complementary to rat alpha 1 and alpha
RT   2 collagen mRNA and their use in studying the regulation of type I
RT   collagen synthesis by 1,25-dihydroxyvitamin D.";
RL   Biochemistry 23:6210-6216(1984).
CC   -!- FUNCTION: Type I collagen is a member of group I collagen
CC       (fibrillar forming collagen).
CC   -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC   -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC       bones. In bones the fibrils are mineralized with calcium
```

```
CC       hydroxyapatite.
CC   -!- PTM: Proline residues at the third position of the tripeptide
CC       repeating unit (G-X-Y) are hydroxylated in some or all of the
CC       chains.
CC   -!- PTM: O-linked glycan consists of a Glc-Gal disaccharide bound to
CC       the oxygen atom of a post-translational added hydroxyl group.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M11432; AAA40832.1; ALT_SEQ.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001007; VWF_C.
DR   ProDom; PD000007; Clg_helix_1; PARTIAL.
DR   PROSITE; PS01208; VWFC_1; PARTIAL.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Pyrrolidone carboxylic acid.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
FT   MOD_RES       9      9       CONVERTED TO AN ALDEHYDE GROUP THAT IS
FT                                INVOLVED IN CROSS-LINKING.
FT   MOD_RES      28     28       HYDROXYLATION (PROBABLE).
FT   MOD_RES      31     31       HYDROXYLATION (PROBABLE).
FT   MOD_RES      34     34       HYDROXYLATION (PROBABLE).
FT   MOD_RES      43     43       HYDROXYLATION (PROBABLE).
FT   MOD_RES      46     46       HYDROXYLATION (PROBABLE).
FT   MOD_RES      49     49       HYDROXYLATION (PROBABLE).
FT   MOD_RES     103    103       HYDROXYLATION (PROBABLE).
FT   CARBOHYD    103    103       O-LINKED (GAL...).
FT   MOD_RES     424    424       HYDROXYLATION (PROBABLE).
FT   MOD_RES     547    547       HYDROXYLATION (PROBABLE).
FT   NON_CONS    567    568
FT   DOMAIN      641    651       MAJOR ANTIGENIC DETERMINANT (OF NEUTRAL
FT                                SALT-EXTRACTED RAT SKIN COLLAGEN).
SQ   SEQUENCE   671 AA;  60615 MW;  9DC3114204AC4918 CRC64;

Query Match             70.1%;  Score 2151;  DB 1;  Length 671;
Best Local Similarity   77.5%;  Pred. No. 9.1e-81;
Matches  382;  Conservative  27;  Mismatches  36;  Indels  48;  Gaps   2;

Qy         1 GSEGPEGVRGEPGPPGPAGAAGPAGPGADGEPGAKGADGAPGIAGAPFPGARGSGPE    60
             ||||||::|:|||||||||||||:|||||||:||::|||||||||:||||||
Db       200 GSBGPQGVRGEPGPPGPAGANGPAGPAGNPGADGQPGAKGANGAPGIAGAPGIAGAPGARGSGPQ  259

Qy        61 GPGGPPGKDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEGKPGARGEPGTGLP  120
             ||||||::|::||||||||||||||||||:||:|||||||||||||||||||||||
Db       260 GPSGAPGKGNSGEPGAPGNKGDTGAKGEPGAGVGPVGPPGPAGEBGRGARGEPGPSGLP  319

Qy       121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEARGPGEAGLPGAKGLT  180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       320 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEARGPGEAGLPGAKGLT  379

Qy       181 GSPGSPGPDGKTGPPGPPGAGEDGRPGPPGPPGGPPGGARGEAGVMGFPGPGKGAAGEPGKAGERGVP  240
             |||||||||||||||||:|||||:||||||||||:::|||||||||||||||||::
Db       380 GSPGSPGPDGKTGPPGAGZBGRPGPAGPPGGPGZBGRPGPAGZBGRPGPAGERGEQGPAGSPGFQGLPGPAGPGEAGKPGAZ  439

Qy       241 GPPGAVGPAGKDGEAGAEGPPGPAGERGEEGPAGSPGFEGLPGPAGPPGEAGKPGEE  300
             |||||||||||||||||||||||:|||||:::||||||:::||||||||||||:::
Db       440 GPPGAVGPAGKDGEAGAQGPAGPAGPAGERGEQGPAGSPGFQGLPGPAGPGEAGKPGZZ  499

Qy       301 GVPGDLGAPGSPGARGBGPGFPGERGVEGPPGPAGPPGADGPGDDGAKGDAGAPGSE  360
             ||||||||||:||||||||||||:|||||||||:::|:||||||::|||||||||||
Db       500 GVPGDLGAPGSPGSGARGBRGFPGBRGVQGPPGPAGPRGNNGAPGNDGAKGDTGAPGAPGSQ  559

Qy       361 GAPGLEGMPGERGAAGLPGPKGDRDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  420
             |||:||||:||:::||:||:
Db       560 GAPGLQGMSGLZ-----------------------------------GPPGPGSSPGZZ  583
```

```
Qy   421  GESGPSPAGPTGARGAPGDRGEPGPPGPAGPAGPPGADGEPGAKGEPGDAGAKGDAGPP  480
          ||| |||||    |||  |  |||||  | |||       |||||||:| |||
Db   584  GPGSASGPAGPR--------GPGSAGSPGKBGLBGLPGPIGPPGPRGRTGBAGPS  631

Qy   481  GPAGPAGPPGPIG  493
          ||||||||||||
Db   632  GPPGPPGPPGPPG  644


RESULT 8
CA11_BOVIN
ID   CA11_BOVIN     STANDARD;      PRT;    779 AA.
AC   P02453;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Collagen alpha 1(I) chain (Fragments).
GN   COL1A1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 1-19.
RX   MEDLINE=72255334; PubMed=4115172;
RA   Rauterberg J., Timpl R., Furthmayr H.;
RT   "Structural characterization of N-terminal antigenic determinants in
RT   calf and human collagen.";
RL   Eur. J. Biochem. 27:231-237(1972).
RN   [2]
RP   SEQUENCE OF 20-145.
RX   MEDLINE=76022320; PubMed=1164916;
RA   Fietzek P.P., Kuehn K.;
RT   "The covalent structure of collagen: amino-acid sequence of the
RT   cyanogen-bromide peptides alpha-1-CB2, alpha-1-CB4 and alpha-1-CB5
RT   from calf-skin collagen.";
RL   Eur. J. Biochem. 52:77-82(1975).
RN   [3]
RP   SEQUENCE OF 146-294.
RX   MEDLINE=73049499; PubMed=4673951;
RA   Fietzek P.P., Wendt P., Kell I., Kuehn K.;
RT   "The covalent structure of collagen: amino acid sequence of alpha-1-
RT   CB3 from calf skin collagen.";
RL   FEBS Lett. 26:74-76(1972).
RN   [4]
RP   SEQUENCE OF 295-562.
RX   MEDLINE=74086118; PubMed=4359390;
RA   Fietzek P.P., Rexrodt F.W., Hopper K.E., Kuehn K.;
RT   "The covalent structure of collagen. 2. The amino-acid sequence of
RT   alpha-1-CB7 from calf-skin collagen.";
RL   Eur. J. Biochem. 38:396-400(1973).
RN   [5]
RP   SEQUENCE OF 563-675.
RX   MEDLINE=73042276; PubMed=4343808;
RA   Wendt P., Mark K.V.D., Rexrodt F., Kuehn K.;
RT   "The covalent structure of collagen. The amino-acid sequence of the
RT   112-residues. Amino-terminal part of peptide alpha-1-CB6 from calf-
RT   skin collagen.";
RL   Eur. J. Biochem. 30:169-183(1972).
RN   [6]
RP   SEQUENCE OF 676-751.
RX   MEDLINE=73042275; PubMed=4343807;
RA   Fietzek P.P., Rexrodt F.W., Wendt P., Stark M., Kuehn K.;
RT   "The covalent structure of collagen. Amino-acid sequence of peptide
RT   alpha-1-CB6-C2.";
RL   Eur. J. Biochem. 30:163-168(1972).
CC   -!- FUNCTION: Type I collagen is a member of group I collagen
CC       (fibrillar forming collagen).
CC   -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC   -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC       bones. In bones the fibrils are mineralized with calcium
CC       hydroxyapatite.


CC   -!- PTM: Proline residues at the third position of the tripeptide
CC       repeating unit (G-X-Y) are hydroxylated in some or all of the
CC       chains. Pro-726 is the only 3-hydroxypro and the only hydroxylated
CC       proline in position X.
CC   -!- PTM: O-linked glycan consists of a Glc-Gal disaccharide bound to
CC       the oxygen atom of a post-translationally added hydroxyl group.
CC   -!- MISCELLANEOUS: The complete chain contains 1052 residues.
DR   PIR; A9119; CGBO1S.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001007; VWF_C.
DR   ProDom; PD000007; Clg_helix_1.
DR   PROSITE; PS01208; VWFC_1; PARTIAL.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Pyrrolidone carboxylic acid.
FT   MOD_RES       1      1     PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES       9      9     CONVERTED TO AN ALDEHYDE GROUP THAT IS
FT                              INVOLVED IN CROSS-LINKING.
FT   MOD_RES     103    103     HYDROXYLATION.
FT   CARBOHYD    103    103     O-LINKED (GAL...).
FT   MOD_RES     115    115     HYDROXYLATION (POTENTIAL).
FT   MOD_RES     124    124     HYDROXYLATION (POTENTIAL).
FT   NON_CONS    145    146
FT   MOD_RES     274    274     HYDROXYLATION (POTENTIAL).
FT   MOD_RES     346    346     HYDROXYLATION (POTENTIAL).
FT   MOD_RES     424    424     HYDROXYLATION (POTENTIAL).
FT   MOD_RES     496    496     HYDROXYLATION (POTENTIAL).
FT   MOD_RES     658    658     HYDROXYLATION (POTENTIAL).
FT   MOD_RES     670    670     HYDROXYLATION (POTENTIAL).
FT   MOD_RES     726    726     HYDROXYLATION.
SQ   SEQUENCE   779 AA;  70346 MW;  E554A7FF084283D1 CRC64;

Query Match             67.6%;  Score 2076.5;  DB 1;  Length 779;
Best Local Similarity   78.0%;  Pred. No. 9.9e-78;
Matches  382;  Conservative  18;  Mismatches  51;  Indels  39;  Gaps  6;

Qy    55  GPSGPEGPGGPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEEGKPGARGEP  114
          ||  ||| :|||  ||
Db    17  GPMGPSGPGRGLPGP-------PGA-----------------PGPQGFQGPP---GEPGEPGASGPM  55

Qy   115  GPTGLPGPPGERGGPGSRGFPGADGVAGPKGYGPAGERGSPGPAGPKGSPGEAGRPGEAGLP  174
          ||||| | | |||| :||||||  :|| ||
Db    56  GPRGPPGPKNGD--------DGEAGKGRPGERGPPGPQGAR------GLPGTAGLP  100

Qy   175  GAKGLTGSPGSPGPDGKTGPPGPAGEDGRPGPGPPGPPGPAGEAGVMGPPGPKGAAGEPGKA  234
          | | :  |||| || |||  |    ||  | | ::||||||||
Db   101  GMKGHRGFSGLLDGAKDAGPAGPKGEPGSPGGENGAPGQMGPRGLPGFPGPKGAAGEPGKA  160

Qy   235  GERGVPGPPGAVGPAGKDGEAGEAGPPGPAGERGEEGPAGSPGFEGLPGPAGPPGEA  294
          ||||| ||||||||| ||||||| |||||||| ||||||||||  :|||||||||
Db   161  GERGVPGPGPAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEA  220

Qy   295  GKPGEEGVPGDLGAPGPSGARGEPGPGPAGPSGARGEPGPGPAGPSGARGERGPPGPAGPPGEA  354
          |||||||||||  :|||  || |||  |  ||  ||  |:|  || |:||||  ||||||
Db   221  GKPGEQGVPGDLGAPGPSGARGERGPPGERGVBGPPGERGVBGPPGRGANGAPGNDGAKGDAGAP  280

Qy   355  GAPGSEGAPGLEGMPGERGAAGLPGKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPA  414
          ||||||| ||| ||  || ||| :| :|| || || || ||| |
Db   281  GAPGSQGAPGLQGMPGERGAAGLPGKDGADGAPGKDGAPGKDGVRGLTGPIGPPGPA  340

Qy   415  GAPGDKGESGPSGPAGPTGARGAPGDRGEPGEPGPPGPAGFAGPPGPAKGEPGDAGAK  474
          |||| ||:|||  ||||  ||| |   :|||  ||||  || ||||||||||| ||||
Db   341  GAPGDKGEAGPSGPA---GTRGAPGDRGEPGEPGPPGPAGFAGPPGPAKGEPGDAGAK  397

Qy   475  GDAGPPGAPGAPGPPGPIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGDAGPP  534
          ||||||||||| |||  || :|||  ||||||||| |  :||  ||||||||  |||||
Db   398  GDAGPPGPAGPPGPIGNVGAPGPKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPP  457

Qy   535  GPPGPAGKEG  544
          |||||||||
Db   458  GPPGPAGKEG  467
```

```
RESULT 9
CA21_BOVIN
ID   CA21_BOVIN     STANDARD;      PRT;    1364 AA.
AC   P02465; 062649;
DT   21-JUL-1986 (Rel. 01, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Collagen alpha 2(I) chain precursor.
GN   COL1A2.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Aorta;
RX   MEDLINE=98290219; PubMed=9628255;
RA   Shirai T., Hattori S., Sakaguchi M., Inoue S., Kimura A., Ebihara T.,
RA   Irie S., Nagai Y., Hori H.;
RT   "The complete cDNA coding sequence for the bovine proalpha2(I) chain
RT   of type I procollagen.";
RL   Matrix Biol. 17:85-88(1998).
RN   [2]
RP   SEQUENCE OF 80-98.
RC   TISSUE=Skin;
RX   MEDLINE=75036115; PubMed=4609475;
RA   Fietzek P.P., Breitkreutz D., Kuehn K.;
RT   "Amino acid sequence of the amino-terminal region of calf skin
RT   collagen.";
RL   Biochim. Biophys. Acta 365:305-310(1974).
RN   [3]
RP   SEQUENCE OF 95-415, AND REVISION.
RC   TISSUE=Skin;
RX   MEDLINE=76091874; PubMed=173531;
RA   Fietzek P.P., Rexrodt F.W.;
RT   "The covalent structure of collagen. The amino-acid sequence of
RT   alpha2-CB4 from calf-skin collagen.";
RL   Eur. J. Biochem. 59:113-118(1975).
RN   [4]
RP   SEQUENCE OF 416-445.
RC   TISSUE=Skin;
RX   MEDLINE=75008198; PubMed=4412529;
RA   Fietzek P.P., Furthmayr H., Kuehn K.;
RT   "Comparative sequence studies on alpha2-CB2 from calf, human, rabbit
RT   and pig-skin collagen.";
RL   Eur. J. Biochem. 47:257-261(1974).
RN   [5]
RP   SEQUENCE OF 446-481.
RC   TISSUE=Skin;
RX   MEDLINE=75059250; PubMed=4435743;
RA   Fietzek P.P., Kuehn K.;
RT   "The covalent structure of collagen: amino acid sequence of the N-
RT   terminal region of alpha2-CB3 from rat skin collagen and alpha2-CB3.5
RT   from calf skin collagen.";
RL   Hoppe-Seyler's Z. Physiol. Chem. 355:647-650(1974).
CC   -!- FUNCTION: Type I collagen is a member of group I collagen
CC       (fibrillar forming collagen).
CC   -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC   -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC       bones. In bones the fibrils are mineralized with calcium
CC       hydroxyapatite.
CC   -!- PTM: Prolines at the third position of the tripeptide repeating
CC       unit (G-X-Y) are hydroxylated in some or all of the chains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB008683; BAA25171.1; -.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD000007; Clg_helix; 3.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Collagen; Signal; Pyrrolidone carboxylic acid.
FT   SIGNAL        1     24
FT   PROPEP       25     79       AMINO-TERMINAL PROPEPTIDE.
FT   CHAIN        80   1100       COLLAGEN ALPHA 2(I) CHAIN.
FT   PROPEP     1101   1364       CARBOXYL-TERMINAL PROPEPTIDE.
FT   MOD_RES      80     80       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES      84     84       CONVERTED TO AN ALDEHYDE GROUP THAT IS
FT                                INVOLVED IN CROSS-LINKING (PROBABLE).
FT   MOD_RES     175    175       HYDROXYLATION (PROBABLE).
FT   MOD_RES     196    196       HYDROXYLATION (PROBABLE).
FT   MOD_RES     262    262       HYDROXYLATION (PROBABLE).
FT   MOD_RES     307    307       HYDROXYLATION (PROBABLE).
FT   MOD_RES     352    352       HYDROXYLATION (PROBABLE).
FT   CONFLICT    157    157       V -> P (IN REF. 3).
FT   CONFLICT    187    187       K -> T (IN REF. 3).
FT   CONFLICT    211    211       T -> K (IN REF. 3).
FT   CONFLICT    298    300       PGA -> AGP (IN REF. 3).
FT   CONFLICT    423    424       AT -> TA (IN REF. 4).
SQ   SEQUENCE   1364 AA;  129064 MW;  5593F4D6B9ED119A CRC64;
```

```
Query Match             63.5%;  Score 1949;  DB 1;  Length 1364;
Best Local Similarity   65.3%;  Pred. No. 1.9e-72;
Matches  355;  Conservative  44;  Mismatches 145;  Indels   0;  Gaps   0;

Qy      1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE 60
            |||||| :||| ||| ||:||::|||||  ||||  ||||:| |||:| |||||||||||
Db    272 GPAGPAGPGEVGLPGLSGPVGPPGPGNPGANGLPGAKGAAGLPGVAGAPGLPGPRGIPGPV 331

Qy     61 GPGGPPGPGKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPPAGEEGKPGARGEPGPTGLP 120
            |||| |:|||||| ||:::||||:|||||| ||| |:||||||||| ||||||||||||
Db    332 GAAGATGARGLVGEPGPAGSKGESGNKGEPGAVGVGQPGPGPSGEEGKRGSTGIGPAGPP 391

Qy    121 GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
            ||||||| ||:||| | :||  |||  ::|||| | :||:|| ||:|||||||| :|
Db    392 GPGLRGNPGSRGLPGADGRAGVMGPAGSRGATGPAGVRGPNGDSGRPGEPGLMGPRGFPP 451

Qy    181 GSPGSPGDGKTGPPGPAGEDGRPGPPGGPGARGEAGVMGFPGPGPKGAAGEPGKAGERGVP 240
            |||| |||:||| |||||||| ||| : |||::| ::||| |:|:||||| :|:
Db    452 GSPGNIGPAGKSGPVGLPGIDGRPGPIGPAGARGEPGNIGFPGPGPKGPSGDPGKAGEKGHA 511

Qy    241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEGPAGSPGFGLPGPAGPPGEAGKPGGE 300
            ||| | |||| || | |:| :||::| :||| |||||| ||::|||||| :|:|||||
Db    512 GLAGARGAPGPDGNNGAQGPPGLQGVQGKGEQGPAGPPGPFQGLPGPAGTAGEAGKPGER 571

Qy    301 GVPGDLGAPGSGARGEPGAGPEPGERGVEGPPGPPGPADGANGRGDGAKGDAGAPGAPGSE 360
            |:|::|||:||| | :||| | ||| |:| |||| |||||:: :||| |:|| |:|| |:|
Db    572 GIPGEFGLPGPAGARGERGPPGESGAAGPTGPIGSRGPGSPPGPGDGNKGEPGVVGAPGTA 631

Qy    361 GAPGLEGMPGERGAAGLPGPKGKGDRGDRAGPKGKADGSPGKDGVRGLTGPIGPPGPAGPGDK 420
            ||| :|| |||||| ||||||:  |: | :| :|| :||| ||:|| : ||| ||:|:|||
Db    632 GPGSPSGLPGERGAAGIPGGKGKGETGLRGDIGSPGRDGARGAPGAIGAPGPANGANGDR 691

Qy    421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGFGPGADGEPGAKGEPGDAGAKGDAGPP 480
            |||||:| |||| |:| ||||| ||||||||||||:::| |:|| ||||: | |||||||
Db    692 GEAGPAGPAGPGPGERGEVGPAGPNGFAGPAGAGQPGAKGQPGAKGRGTKGPKGENGPV 751

Qy    481 GPAGPAGPGPGIGDVGAPGAKGARGSAGPPGATGFPGAAGRVGPGPGSGDAGPGGPGPA 540
            ||||||||| ||| :|: :||| | || |||| :|:::| ||:| |: | ||| ||||
Db    752 GPTGPVGAAGPSGPNGPPGPAGSRGDGGPPGATGFPGAAGRTGPPGPSGISGPPGGPGPA 811

Qy    541 GKEG 544
          ||||
Db    812 GKEG 815
```

```
RESULT 10
CA13
ID   CA13_CHICK     STANDARD;      PRT;  1262 AA.
AC   P12105; P79758; P79759; Q90794; Q92029;
DT   01-OCT-1989 (Rel. 12, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Collagen alpha 1(III) chain precursor (Fragments).
GN   COL3A1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE OF 1-886 FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=94266842; PubMed=8206952;
RA   Nah H.-D., Niu Z., Adams S.L.;
RT   "An alternative transcript of the chick type III collagen gene that
RT   does not encode type III collagen.";
RL   J. Biol. Chem. 269:16443-16448(1994).
RN   [2]
RP   SEQUENCE OF 29-96, 332-397, 431-484, 503-535 AND 869-976 FROM N.A.
RX   MEDLINE=84270696; PubMed=654777O;
RA   Yamada Y., Liau G., Mudryj M., Obici S., de Crombrugghe B.;
RT   "Conservation of the sizes for one but not another class of exons in
RT   two chick collagen genes.";
RL   Nature 310:333-337(1984).
RN   [3]
RP   SEQUENCE OF 977-1262 FROM N.A.
RX   MEDLINE=83220816; PubMed=6856474;
RA   Yamada Y., Kuhn K., de Crombrugghe B.;
RT   "A conserved nucleotide sequence, coding for a segment of the C-
RT   propeptide, is found at the same location in different collagen
RT   genes.";
RL   Nucleic Acids Res. 11:2733-2744(1983).
CC   -!- FUNCTION: Collagen type III occurs in most soft connective tissues
CC       along with type I collagen.
CC   -!- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are
CC       linked to each other by interchain disulfide bonds. Trimers are
CC       also cross-linked via hydroxylysines.
CC   -!- PTM: Prolines at the third position of the tripeptide repeating
CC       unit (G-X-Y) are hydroxylated in some or all of the chains.
CC   -!- SIMILARITY: Contains 1 VWFC domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U07973; AAA83407.1; -.
DR   EMBL; X00822; CAB52686.1; -.
DR   EMBL; X00823; CAB52686.1; JOINED.
DR   EMBL; X00826; CAA25397.1; ALT_SEQ.
DR   EMBL; X00825; CAA25397.1; JOINED.
DR   EMBL; X00827; CAA25398.1; -.
DR   EMBL; X00828; CAA25399.1; -.
DR   EMBL; X00830; CAA25401.1; -.
DR   EMBL; X00831; CAA25402.1; -.
DR   EMBL; K02302; AAD15299.1; -.
DR   EMBL; K02301; AAD15298.1; -.
DR   EMBL; M36662; AAA18519.1; ALT_SEQ.
DR   PIR; A05269; A05269.
DR   PIR; I50694; I50694.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR002181; Fibrinogen_C.
DR   InterPro; IPR001007; VWF_C.
DR   ProDom; PD000007; Clg_helix; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM000038; COLFI; 1.
DR   SMART; SM002014; VWC; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal.
FT   SIGNAL        1     23
FT   PROPEP       24    144       AMINO-TERMINAL PROPEPTIDE (BY
FT                                SIMILARITY).
FT   CHAIN       145   1003       COLLAGEN ALPHA 1(III) CHAIN.
FT   PROPEP     1004   1262       CARBOXYL-TERMINAL PROPEPTIDE (BY
FT                                SIMILARITY).
FT   DOMAIN       29     88       VWFC.
FT   DOMAIN      145    164       NONHELICAL REGION (N-TERMINAL) (BY
FT                                SIMILARITY).
FT   DOMAIN      165    994       TRIPLE-HELICAL REGION (BY SIMILARITY).
FT   DOMAIN      995   1003       NONHELICAL REGION (C-TERMINAL) (BY
FT                                SIMILARITY).
FT   NON_CONS    886    887
FT   NON_CONS    922    923
FT   DISULFID    994    994       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    995    995       INTERCHAIN (BY SIMILARITY).
FT   MOD_RES     262    262       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES     283    283       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES     859    859       HYDROXYLATION (BY SIMILARITY).
FT   CARBOHYD   1163   1163       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT     96     96       E -> K (IN REF. 2).
FT   CONFLICT   1132   1132       F -> S (IN REF. 3).
SQ   SEQUENCE  1262 AA;  121249 MW;  96ABE7B2E9DEB43D CRC64;
```

```
Query Match             63.0%;
Best Local Similarity   63.8%;
Matches  353;  Conservative  31;                Score 1934.5;  DB 1;  Length 1262;
               Mismatches 160;  Indels   9;  Gaps   9;
Pred. No. 7e-72;

Qy      1  GSEGPEGVRGEPGPPGPPGAGAAGPAGDGADGEPGADGEPGADGAPGIAGAPGFPGARGPSGPE      60
           :.|  ||||||  | ||.||||||||  ||  ||||| |||||||||||| ||| |||||||:|
Db    359  GASGNPGERGEPGPQGQAGPGGPGPGPPGRAGSPGGGKGEMGPSGLPGGPGPGGRGLPGGPP     418

Qy     61  GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGPAGEBEGKPGGARGEPGPTGLP     120
           |||||| |.:|  |||||||:  || ||| ||||||| |||||||| || || ||||| | |
Db    419  GTSGNPGAKGTPGEPGKNGAKGDPGPKGERGENGTPGARGPPGEEGKRGANGEPGQNGVP     478

Qy    121  GPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGERGSPGEAGRPGEAGLPGAKGLT     180
           |||||| |:| |  ||.:.  |||| ||:|| |||||||  :::.||| |||  ||||:|
Db    479  GTPGERGSPGFRGLPGSNGLPGEKGPAGERGSPGPPGPGGRGLPGMRGLP     538

Qy    181  GSPGSPGPDGKTGPPGPAGEDGRPGPPGGARGEAGVMGFPGPKGAAGEPGKAGERGVP     240
           ||||||| |  :||||| |:  |||:|| || || || |||||||| ||||||||||:
Db    539  GIPGSPGSDGKPGPPGNQGPGNQGEPGRSGPPGPAGPRQPGVWGFPGPKGNEGAPGKNGERGPG     598

Qy    241  GPPGAVGPAGKDGDEAGAGEPGPPGPAGPAGERGEEGPAGSPGFEGLPGAGPPGPAGEAGKPGSE     300
           |||  || ||  |||||||||||||||||||:||||||||||||||||||| || ||  || |||
Db    599  GPPGTPGPAGKNGDVGLPGPPGADRGEEPGPSGSPGLQGLPGGPGPGPAGENGKPGEP     658

Qy    301  GVPGDLGAPGPSGARGEBGPGPGERGCVEGPPGPAGPPGADGAPGDDGAKGDAGAPGAPGSE     360
           |||| || || ||| || ||| ||| :::.||||||| |||:.|| |||||  |||||||
Db    659  GPKGDIGGPGFPGPKGENGIPGERGPQGPPGPGTGARGGPGAGSEGAKGPPGPGAPGGT     718

Qy    361  GAPGLEGMPGERGAAGLPGPGVRGDRGDAGPKGDRGDSPGGPGVRGLTGPIGPGPAGPGDK     420
           |||| || || |||||||||| |||||||| || |||| || || || ||||||||||:|
Db    719  GLPGLQGMPGERGASGSPGPGKGDKGEPGGKGADGLPGARGERGNVGPIGPPGPAGPPGDK     778

Qy    421  GESGPSGPAGPTGARGAPGAPGEPGPPGADGAPGAPGPAGPDAGAPGDAGAKGDAGPP     480
           |||||||||| || ||| || |||||||||||| |||| | ||| |||:.|||:|||
Db    779  GETGPAGAPGPAGSRGGPGERGEQGLPGPAGFPGAPGFPGAPGQNGEPGGKGERGQPGLRGEAGPP     838

Qy    481  GPAGAGPPGPIGDVGAPGAKGARGSAGPGAGTGFPGAAGRVGPPGPGD--------A     531
           || | |||| |  |||| |||| || | |||:||||| | | ||||| |
Db    839  GAAGPQGGPGAPGPPGPQGVKGERGSPGGPGAAGPFPGARGPPGPGNNGDRGESGPPGVP     898
```

```
Qy    532 GPPGPGPGAGKEG 544
          ||||||||| |
Db    899 GPPGHGPGAGNNG 911


RESULT 11
CA21_RAT
ID   CA21_RAT      STANDARD;      PRT;   1372 AA.
AC   P02466; Q9R1E8;
DT   21-JUL-1986 (Rel. 01, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Collagen alpha 2(I) chain precursor.
GN   COL1A2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Guenther D., Seibold S., Marx M.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 86-98.
RC   TISSUE=Skin;
RX   MEDLINE=67162268; PubMed=5337886;
RA   Kang A.H., Bornstein P., Piez K.A.;
RT   "The amino acid sequence of peptides from the cross-linking region of
RT   rat skin collagen.";
RL   Biochemistry 6:788-795(1967).
RN   [3]
RP   SEQUENCE OF 99-102.
RC   TISSUE=Skin;
RX   MEDLINE=69206881; PubMed=5785232;
RA   Fietzek P.P., Piez K.A.;
RT   "Isolation and characterization of the cyanogen bromide peptides from
RT   the alpha 2 chain of rat skin collagen.";
RL   Biochemistry 8:2129-2133(1969).
RN   [4]
RP   SEQUENCE OF 102-144.
RC   TISSUE=Skin;
RX   MEDLINE=73049496; PubMed=4636752;
RA   Fietzek P.P., Kell I., Kuehn K.;
RT   "The covalent structure of collagen. Amino acid sequence of the N-
RT   terminal region of alpha 2-CB4 from calf and rat skin collagen.";
RL   FEBS Lett. 26:66-68(1972).
RN   [5]
RP   SEQUENCE OF 423-452.
RC   TISSUE=Skin;
RX   MEDLINE=71115216; PubMed=5544653;
RA   Highberger J.H., Kang A.H., Gross J.;
RT   "Comparative studies on the amino acid sequence of the alpha 2-CB2
RT   peptides from chick and rat skin collagens.";
RL   Biochemistry 10:610-616(1971).
RN   [6]
RP   SEQUENCE OF 453-501.
RC   TISSUE=Skin;
RX   MEDLINE=75059250; PubMed=4435743;
RA   Fietzek P.P., Kuehn K.;
RT   "The covalent structure of collagen: amino acid sequence of the N-
RT   terminal region of alpha2-CB3 from rat skin collagen and alpha2-CB3.5
RT   from calf skin collagen.";
RL   Hoppe-Seyler's Z. Physiol. Chem. 355:647-650(1974).
RN   [7]
RP   SEQUENCE OF 791-836.
RC   TISSUE=Skin;
RX   MEDLINE=74055004; PubMed=4763308;
RA   Fietzek P.P., Kuehn K.;
RT   "The covalent structure of collagen: amino acid sequence of the N-
RT   terminal region of alpha 2-CB5 from rat skin collagen.";
RL   FEBS Lett. 36:289-291(1973).
RN   [8]
RP   ORDER OF CNBR PEPTIDES.
RX   MEDLINE=70181852; PubMed=5443712;
RA   Vuust J., Lane J.M., Fietzek P.P., Miller E.J., Piez K.A.;
RT   "The order of the CNBr peptides from the alpha 2 chain of collagen.";
RL   Biochem. Biophys. Res. Commun. 38:703-708(1970).
CC   -!- FUNCTION: Type I collagen is a member of group I collagen
CC       (fibrillar forming collagen).
CC   -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC   -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC       bones. In bones the fibrils are mineralized with calcium
CC       hydroxyapatite.
CC   -!- PTM: Prolines at the third position of the tripeptide repeating
CC       unit (G-X-Y) are hydroxylated in some or all of the chains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF121217; AAD41775.1; -.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD000007; Clg_helix; 4.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal; Pyrrolidone carboxylic acid.
FT   SIGNAL        1     24       POTENTIAL.
FT   PROPEP       25     85       AMINO-TERMINAL PROPEPTIDE
FT                                (BY SIMILARITY).
FT   CHAIN        86   1108       COLLAGEN ALPHA 2(I) CHAIN.
FT   PROPEP     1109   1372       CARBOXYL-TERMINAL PROPEPTIDE
FT                                (BY SIMILARITY).
FT   SITE        783    785       CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE        828    830       CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE       1011   1013       CELL ATTACHMENT SITE (POTENTIAL).
FT   MOD_RES      86     86       PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
FT   MOD_RES      90     90       CONVERTED TO AN ALDEHYDE GROUP THAT IS
FT                                INVOLVED IN CROSS-LINKING.
FT   CARBOHYD   1273   1273       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT    132    132       T -> P (IN REF. 4).
FT   CONFLICT    137    137       S -> P (IN REF. 4).
FT   CONFLICT    145    422       MISSING (IN REF. 4).
FT   CONFLICT    431    432       ST -> TS (IN REF. 5).
FT   CONFLICT    494    494       E -> Z (IN REF. 6).
FT   CONFLICT    497    497       N -> A (IN REF. 6).
FT   CONFLICT    502    790       MISSING (IN REF. 6).
FT   CONFLICT    825    825       R -> K (IN REF. 7).
SQ   SEQUENCE   1372 AA;  129564 MW;  B06971A8DB20A72 CRC64;

Query Match             63.0%;  Score 1934;  DB 1;  Length 1372;
Best Local Similarity   64.7%;  Pred. No. 7.8e-72;
Matches  352;  Conservative  44;  Mismatches 148;  Indels    0;  Gaps    0;

Qy      1 GSEGPEGVRGEPGPPGPAGAAGPAGADPAGDPGAKGADGAPGIAGAPGFPGARGPSGPE 60
          ||||||: |:|| ||| | |:  : |   |  || |:||:|| |: || : ||||||||
Db    280 GPAGPAGPRGEAGLPGLSGPVGPGPGNPGANGLTGAKGATGLPGVAGAPGLPGPRGIPGPV 339

Qy     61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGVGVEGPPGPGPAGEEGKPGARGEPGPTGLP 120
          | |: ||||  | |   || |  |  | || :  | ||||  | | |||||: ||| |||
Db    340 GAAGATGPKGRGLVGEPGPAGSKGETGNKGEPGGSAGAQGPPGPSGEEGKRGSPGBPGSAGPA 399

Qy    121 GPPGBERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
          ||  ||:||: |||| |||| | | | ||||: || |: ||: : ::|| |||  | ::
Db    400 GPPGLRGSPGSRGLPGADGRAGVMGPPGNRGSTGPAGVRGPNGDAGRPGEPGLMGPRGLP 459

Qy    181 GSPGSPGDGKTGPPGPAGEDGRPGPPGPPGPGARGEAGVMGPGPGPKGAAGEPGKAGERGVP 240
```

```
Db       |||:.|| |||| |    ||||:|||||||| |||| :||||||| ||||| .:|:|||||:||
     460 GSPGNVGPAGKEGVGLPGIDGRPGPIGPAGPRGEAGNIGFPGPKGPSGDPGKPGEKGHP 519

Qy       |||||  ||  | ||||||| |||||| |||| ||| |||||||||| ||| |||||||||
     241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGERGEGPAGSPGFEGLPGPAGPPGEAGKPGEE 300

Db       |||:::| |||||||||||||||:||||||||||||||||:|||||||||||||||||||
     520 GLAGARGAPGPDGNNGAQGPGPGPGOVGGKGEQCPAGPPGFGQLPGFPSGTAGEVGKPGER 579

Qy       |||: :| |:||| ||| ||||||| ||||||||||||||||||||||||||||||||||
     301 GVPGDLGAPGSGAGERGEFGPGPGERGVEGPGPAGPPGADGAPGGDGAKGDAGAPGAPGSE 360

Db       |:.:||| ||||||||| :||||||| |||||||||||||| ||||.||||||||||||
     580 GLPGBFGLPGPAGPRGERGPPGESGAAGPSGPIGIRGPSGAPGPDGNKGEAGAVGAPGSA 639

Qy       || :|||||||||||  |: ||||| || :: || |:|:|||||||||||||||||||||
     361 GAPGLEGMPGERGAAGLPGPKGDRDGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK 420

Db       ||: |||||||||||| |:: :|:|:|||||||||||||:||:| |||||||||||||
     640 GASGPGGLPGERGAAGIPGGKGEKGETGLRGEIGNPGRDGARGAPGAIGAPGPAGASGDR 699

Qy       |||:|||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
     421 GESGPSGPAGPTGARGAPGDRGEPGPGPAGPPGADGEPGAKGEPGDAGAKGDGAGPP 480

Db       |||: :||:| :|||||||:| ||||| |  |:||:| ||:||||||||||||||
     700 GEAGAAGPSGPAGPRGSPGERGEVGPAGPNGFAGPAGSAGQPGAKGAKGEKGTKGPKGENGIV 759

Qy       ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
     481 GPAGPAGPPGPIGDVGAGAKGARGSAGPPGATGPGAAGRVGPPGPSGDAGPPGPPGPA 540

Db       |::.|:  |||||    |:|| ||||||||||||||||||||||||||||||||||||||
     760 GPTGPVGAAGPSGPNGPPGPAGSRGDGGPPGMTGFPGAAGRTGPPGPSGITGPPGPPGAA 819

Qy   541 GKEG 544

Db   820 GKEG 823
```

RESULT 12

ID   CA25_HUMAN   STANDARD;   PRT;   1496 AA.
AC   P05997;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Collagen alpha 2 (V) chain precursor.
GN   COL5A2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-463 FROM N.A.
RX   MEDLINE=89123368; PubMed=2914927;
RA   Woodbury D., Benson-Chanda V., Ramirez F.;
RT   "Amino-terminal propeptide of human pro-alpha 2(V) collagen conforms
RT   to the structural criteria of a fibrillar procollagen molecule.";
RL   J. Biol. Chem. 264:2735-2738(1989).
RN   [2]
RP   SEQUENCE OF 398-1496 FROM N.A.
RX   MEDLINE=87146331; PubMed=3029669;
RA   Weil D., Bernard M.P., Gargano S., Ramirez F.;
RT   "The pro alpha 2(V) collagen gene is evolutionarily related to the
RT   major fibrillar-forming collagens.";
RL   Nucleic Acids Res. 15:181-198(1987).
RN   [3]
RP   SEQUENCE OF 1227-1496 FROM N.A.
RX   MEDLINE=85289337; PubMed=2411731;
RA   Myers J.C., Loidl H.R., Seyer J.M., Dion A.S.;
RT   "Complete primary structure of the human alpha 2 type V procollagen
RT   COOH-terminal propeptide.";
RL   J. Biol. Chem. 260:11216-11222(1985).
RN   [4]
RP   SEQUENCE OF 1449-1496 FROM N.A.
RX   MEDLINE=89138450; PubMed=3224983;
RA   Tsipouras P., Schwartz R.C., Liddell A.C., Salkeld C.S., Weil D.,
RA   Ramirez F.;
RT   "Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2,
RT   located on the long arm of human chromosome 2.";
RL   Genomics 3:275-277(1988).
RN   [5]
RP   SEQUENCE OF 208-227.
RC   TISSUE=Placenta;
RX   MEDLINE=92239022; PubMed=1571108;
RA   Mann K.;
RT   "Isolation of the alpha 3-chain of human type V collagen and
RT   characterization by partial sequencing.";
RL   Biol. Chem. Hoppe-Seyler 373:69-75(1992).
RN   [6]
RP   SEQUENCE OF 288-297 AND 606-617.
RC   TISSUE=Bone;
RX   MEDLINE=94237164; PubMed=8181482;
RA   Moradi-Ameli M., Rousseau J.C., Kleman J.P., Champliaud M.F.,
RA   Boutillon M.M., Bernillon J., Wallach J.M., van der Rest M.;
RT   "Diversity in the processing events at the N-terminus of type-V
RT   collagen.";
RL   Eur. J. Biochem. 221:987-995(1994).
RN   [7]
RP   DISEASE.
RX   MEDLINE=98087576; PubMed=9425221;
RA   Michalickova K., Susic M., Willing M.C., Wenstrup R.J., Cole W.G.;
RT   "Mutations of the alpha2(V) chain of type V collagen impair matrix
RT   assembly and produce Ehlers-Danlos syndrome type I.";
RL   Hum. Mol. Genet. 7:249-255(1998).
RN   [8]
RP   VARIANT EDS-II ARG-960.
RX   MEDLINE=98455031; PubMed=9783710;
RA   Richards A.J., Martin S., Nicholls A.C., Harrison J.B., Pope F.M.,
RA   Burrows N.P.;
RT   "A single base mutation in COL5A2 causes Ehlers-Danlos syndrome type
RT   II.";
RL   J. Med. Genet. 35:846-848(1998).
CC   -!- FUNCTION: Type V collagen is a member of group I collagen
CC   (fibrillar forming collagen). It is a minor connective tissue
CC   component of nearly ubiquitous distribution. Type V collagen binds
CC   to DNA, heparan sulfate, thrombospondin, heparin, and insulin.
CC   -!- SUBUNIT: Trimers of two alpha 1(V) and one alpha 2(V) chains in
CC   most tissues and trimers of one alpha 1(V), one alpha 2(V), and
CC   one alpha 3(V) chains in placenta.
CC   -!- PTM: Prolines at the third position of the tripeptide repeating
CC   unit (G-x-Y) are hydroxylated in some or all of the chains.
CC   -!- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome
CC   type I (EDS-I) [MIM:130000]; also known as Ehlers-Danlos syndrome
CC   gravis. EDS-I is an autosomal dominant connective-tissue disorder
CC   characterized by loose-jointedness and fragile, velvety,
CC   stretchable, bruisable skin that heals with peculiar 'cigarette-
CC   paper' scars. Inheritance is autosomal dominant.
CC   -!- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome
CC   type II (EDS-II) [MIM:130010]; also known as Ehlers-Danlos
CC   syndrome mitis. Inheritance is autosomal dominant.
CC   -!- SIMILARITY: Contains 1 VWFC domain.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; J04478; AAA51859.1; -.
DR   EMBL; X04758; CAA28454.1; -.
DR   EMBL; M11718; AAA52058.1; -.
DR   PIR; A31427; CGHU2V.
DR   PDB; 1A9A; 18-NOV-98.
DR   Genew; HGNC:2210; COL5A2.
DR   MIM; 120190; -.
DR   MIM; 130000; -.
DR   MIM; 130010; -.
DR   GO; GO:0005588; C:collagen type V; TAS.
DR   GO; GO:0008151; P:cell growth and/or maintenance; TAS.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR002181; Fibrinogen_C.
DR   InterPro; IPR001007; VWF_C.

```
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF00093; vwc; 1.
DR   ProDom; PD000007; Clg_helix; 5.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal; Ehlers-Danlos syndrome;
KW   Disease mutation; 3D-structure;
FT   SIGNAL        1     26
FT   CHAIN        27   1226       COLLAGEN ALPHA 2 (V) CHAIN.
FT   PROPEP     1227   1496       CARBOXYL-TERMINAL PROPEPTIDE.
FT   DOMAIN       39     97       VWFC.
FT   MOD_RES     290    290       HYDROXYLATION.
FT   MOD_RES     293    293       HYDROXYLATION.
FT   MOD_RES     296    296       HYDROXYLATION.
FT   MOD_RES     608    608       HYDROXYLATION.
FT   MOD_RES     614    614       HYDROXYLATION.
FT   VARIANT     960    960       G -> R (in EDS-II).
FT                               /FTId=VAR_013588.
FT   CONFLICT    292    292       A -> P (IN REF. 6).
FT   CONFLICT   1418   1418       K -> T (IN REF. 3).
FT   CONFLICT   1438   1438       F -> S (IN REF. 3).
FT   CONFLICT   1460   1460       E -> Q (IN REF. 4).
FT   CONFLICT   1496   1496       V -> A (IN REF. 4).
SQ   SEQUENCE  1496 AA;  144720 MW;  82827C17A8644F5A CRC64;

Query Match             62.6%;
Best Local Similarity   64.0%;
Matches  348;  Conservative  40;  Mismatches  153;  Indels  3;  Gaps  1;
                                   Score 1921.5;  DB 1;  Length 1496;
                                   Pred. No. 2.6e-71;

Qy      1 GSEGPEGVRGEPGPGPGPGPAGAAGPGPGDGDGPDADGDGEPGAKGADGAGAGAPGIGPGARGPSGPE 60
            ||:.|||||||:||||      |.|||      :|         |:||||||::||.|:::|:||:||||:||
Db    396 GPEGPQGQRGETGPPGPVGSPGLPGAIGDGTPGPGKGPTGSGP---PGSAGPGPGSP 452

Qy     61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPGPGPAGEGKPGARGEPGPTGLP 120
            ||.|||||.::|:||:||||.|||  :|||:.|.:|||||:||:||:|.:||||.|||
Db    453 GPQGSTGPQGNSGLPGDPGFKGEBAGPGKGEKGEPGPHGIQGPIGPPGEBKGRGRGDDPGTLGPP 512

Qy    121 GPPGERGRGGFPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
            ||.|||||.||:||:||:|||.:|.:|||:||.:|.|||||.:|.:||:|||:||:|||||||
Db    513 GPVGERGAPGNRGFFGSDGLPGPKGAQGERGPVGSSGPKGSQGDPGRPGEPGLPGARGLT 572

Qy    181 GSPSPGPGDGKTGPPGPAGEDGRPGPPGPPGPGGSIGIKGQPGTMGLPGPKGSNGDPGKPGEAGNP 240
            |.:|.|||||.:||:|||:||.||.||||||||.:||.:||||.||:||||||.|||||:||:.|:
Db    573 GNPGVQGPPGKLGPLGAPGEDGRPGPPGSIGIKGQPGTMGLPGPKGSNGDPGKPGEAGNP 632

Qy    241 GPPGAVGPAGKDGEAGAEGPGPVGPAGPAGERGEBGPAGSPGFEGLPGAPGPPGEAGKPGEE 300
            ||.|.:||.|||.:|||:||:||.|||||:||.||:|||.|.|.||||.|||||:|||:||.
Db    633 GVPGDLGAPGSGARGEPGPGERGVEGPPGPAGPPGPGADDGAKGDAGAPAGPGSE 692

Qy    301 GVPGDLGAPGSGARGEPGPGERGVEGPPGPAGPPGPGADGDGAKGDAGAPAGPGSE 360
            ||:|||||.|:|.||:|||:||:||.:|||.||.||||:|:||:|||||||:|||.
Db    693 GVPGGPAVGPLGPRGERGNPGRGEPGEPGITGLPGEKGMGAGGHGPDGPGKGSPGSGTPGDT 752

Qy    361 GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPPAGAPGDK 420
            ||.||:|:|||||.:|||||||.||.||.||||.||.:|:.|.|.|||||.||.|:.||||
Db    753 GPPGLQGMPGERGIAGTPGKGDRGGIGEKGAEGTAGNDGAGGLPGPLGPPGPAGLLGEK 812

Qy    421 GESGPSGPAGPTGARGAPGDRGEPGPPGPAGPAGPGADEGPGADAKGEPGDAGAKGDAGPP 480
            |:|||.|||.||.|||.|||:||.|||||.||:||||.:|||.|.|||.|||||.||||.|
Db    813 GEPGPRGLVGPPGSRGNPGSRGENGPTGAVGFAGPQGSDGQPGVKGEPGEPGGQKGDAGSP 872

Qy    481 GPAGPAGPPGPGIGDVGAPGANGAGARGSAGPPGATGPGAAGRVGPGPGPSGDAGPPGPPGPA 540
            ||.|||||.|.|:.:|.:|||.|:|.:||.|:||||:|||.:||:||.||:|.||||||.|||.
Db    873 GPQGLAGSPGPHGPNGVPGLKGGRGTQGPPGATGPPGSSAGRVGPPGPAGPGAPAGPLGPEP 932

Qy    541 GKEG 544
            ||||
Db    933 GKEG 936
```

```
RESULT 13
CA21_MOUSE
ID   CA21_MOUSE     STANDARD;     PRT;   1372 AA.
AC   Q01149;
DT   01-APR-1993 (Rel. 25, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Collagen alpha 2(I) chain precursor.
GN   COL1A2 OR COLA2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Calvaria;
RX   MEDLINE=92372043; PubMed=1505972;
RA   Phillips C.L., Morgan A.L., Lever L.W., Wenstrup R.J.;
RT   "Sequence analysis of a full-length cDNA for the murine pro alpha
RT   2(I) collagen chain: comparison of the derived primary structure with
RT   human pro alpha 2(I) collagen.";
RL   Genomics 13:1345-1346(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57Bl/6J; TISSUE=Mammary gland;
RX   MEDLINE=22388257; PubMed=2477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez J., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 1-110 FROM N.A.
RC   TISSUE=Calvaria;
RX   MEDLINE=92084969; PubMed=1748823;
RA   Phillips C.L., Lever L.W., Pinnell S.R., Quarles L.D.,
RA   Wenstrup R.J.;
RT   "Construction of a full-length murine pro alpha 2(I) collagen cDNA by
RT   the polymerase chain reaction.";
RL   J. Invest. Dermatol. 97:980-984(1991).
RN   [4]
RP   SEQUENCE OF 1-23 FROM N.A.
RX   MEDLINE=87289650; PubMed=3039494;
RA   Rossi P., de Crombrugghe B.;
RT   "identification of a cell-specific transcriptional enhancer in the
RT   first intron of the mouse alpha 2 (type I) collagen gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:5590-5594(1987).
CC   -!- FUNCTION: Type I collagen is a member of group I collagen
CC       (fibrillar forming collagen).
CC   -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC   -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC       bones. In bones the fibrils are mineralized with calcium
CC       hydroxyapatite.
CC   -!- PTM: Prolines at the third position of the tripeptide repeating
CC       unit (G-X-Y) are hydroxylated in some or all of the chains.
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X58251; CAA41205.1; -.
DR   EMBL; BC007158; AAH07158.1; -.
DR   EMBL; BC042503; AAH42503.2; -.
DR   EMBL; K01832; AAA37331.1; -.
DR   PIR; A43291; A43291.
DR   MGD; MGI:88468; Col1a2.
DR   InterPro; IPR008161; C1g_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD000007; C1g_helix; 6.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal; Pyrrolidone carboxylic acid.
FT   SIGNAL     1     22       POTENTIAL.
FT   PROPEP    23     85       AMINO-TERMINAL PROPEPTIDE
FT                             (BY SIMILARITY).
FT   CHAIN     86   1108       COLLAGEN ALPHA 2(I) CHAIN.
FT   PROPEP  1109   1372       CARBOXYL-TERMINAL PROPEPTIDE
FT                             (BY SIMILARITY).
FT   MOD_RES   86     86       PYRROLIDONE CARBOXYLIC ACID (BY
FT                             SIMILARITY).
FT   MOD_RES   90     90       CONVERTED TO AN ALDEHYDE GROUP THAT IS
FT                             INVOLVED IN CROSS-LINKING
FT                             (BY SIMILARITY).
FT   CARBOHYD  1273  1273      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT    15    15      V -> A (IN REF. 4).
FT   CONFLICT  1167  1167      R -> TT (IN REF. 1).
SQ   SEQUENCE  1372 AA;  129557 MW;  0D17DF5D6C1452D1 CRC64;
```

```
Query Match             62.4%;  Score 1915;  DB 1;  Length 1372;
Best Local Similarity   64.2%;  Pred. No. 4.5e-71;
Matches  349;  Conservative  45;  Mismatches 150;  Indels  0;  Gaps  0;

Qy      1 GSEGPEGVRGEPPGPPGPAGAAGPAGPDGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE  60
          |||:||||| |||:|  |  || |:  :|||:|||   :  |||:||:||||  || |
Db    280 GPAGPAGPRGEVGLPGLSGPVGPPGNPGTNGLTGAKGATGLPGVAGAPGLPGPRGIPGPA 339

Qy     61 GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGPVGVEGPPGAGERGKPGARGEPGPTGLP 120
          |||||||:|||  |||| :|||||  |||||||||||||||||| :||  ::||||||||:
Db    340 GAAGATGARGLVGEPGPAGSKGESGNKGEPGSVGAQPPGPSGEEGKRGSPGEAGSAGPA 399

Qy    121 GPPGERGGPSGRGFPGADGVAGPKGPAGERGSSPGAPGKGSPGEAGRPGEAGLPGAKGLT 180
          |||||||||||||  || || |||||  |||||  |||:||  ||| :  |||||:|||
Db    400 GPPGLRGSPGSRGLPGADGRAGVMGPPGNRGSTGPAGIRGPNGDAGRPGEPGLMGPRGLP 459

Qy    181 GSPGSPGPDGKTGPPGPAGEDGRPGPPGPPGARGEAGVMGFPGPKGAAGEPGKAGERGVP 240
          ||:||||  ||| |||||| |||||||:|:||||:||||| ||:|||| :|:|||:|||||
Db    460 GSPGNVGPSGKEGPVGLPGIDGRPGPIGPAGPRGEAGNIGFPGPKGPGDPGKPGERGHP 519

Qy    241 GPPGAVGPAGKDGEAGAEGPPGPPGPAGERGEEGPAGSPGFEGLPGAPGPGEAGKPGEE 300
          |||| || || ||| || |:  |||:  || |||||| ||||:||  ||  ||||||| |
Db    520 GLAGARGAPGPDCNNGAAQGPPGPQGVQGGKGEQGPAGPPGFQGLPGPSGTTGEVGKPGER 579

Qy    301 GVPGDLGAPGPSGARGEPGFPGSRGVEGPPGPAGPPGADGAPGDDGAKGDAGAPGSE 360
          |:||:||  :|||  ||||||| |:|| ||| ||| | :|||:  || ::||  ||
Db    580 GLPGEFGLPGPAGPRGERGTPGEBGAAGPSGPIGSRGPSGPGPDGNKGEAGAVGAPGSA 639

Qy    361 GAPGLEMPGERGAAGLPGPKGDRGADGPKGADGSPGKDGVRGLTGPIGPGPGAGAPGDK 420
          || ||  ||||| ||| :|  :|||::|:| | || | |:| ::||   ||:| ||||
Db    640 GASGPGGLPGERGAAGIPGGKGEKGETLGRGDTGNTGRDGARGIPGAVGAPGPAGASGDR 699

Qy    421 GESGPSGPAGPTCARGAPGDRGEPGPPGPAGFAGPGPGADGPGEAKGEPGDAGAKGDAGPP 480
```

```
Db    700 GEAGAAGPSGPAGPRGSPGERGEVGPAGPNGFAGPAGAAGQPGAKGEKGTKGPKGENGIV 759
          ||:| | :||:||||    |:| |||||  |   |   || || |:   :|||    |
Qy    481 GPAGPAGPPGPICDVGAPGAKGARGSAGPGPGATGTFPGAAGRVGPPGPSGDAGPPGPGPGPA 540
          |||  | ||||   |   |  | | ||   |||   |||  |:|  |  |  |||||||||
Db    760 GPTGSVGAAGPSGPNGPPGPVGSRGDGGPPGMTGFPGAAGRTGPPGPSGIAGPPGPGPGAA 819
```

```
Qy    541 GKEG 544
          ||||
Db    820 GKEG 823
```

```
RESULT 14
CA13_BOVIN
ID   CA13_BOVIN    STANDARD;    PRT;   1049 AA.
AC   P04258;
DT   20-MAR-1987 (Rel. 04, Created)
DT   20-MAR-1987 (Rel. 04, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Collagen alpha 1(III) chain.
GN   COL3A1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 1-242.
RX   MEDLINE=80026026; PubMed=488906;
RA   Fietzek P.P., Allmann H., Rauterberg J., Henkel W., Wachter E.,
RA   Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. I. The amino
RT   acid sequence of the amino terminal region of the alpha 1(III) chain
RT   (positions 1-222)."
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:809-820(1979).
RN   [2]
RP   SEQUENCE OF 243-422.
RX   MEDLINE=80026027; PubMed=488907;
RA   Dewes H., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. II. The amino
RT   acid sequence of the cyanogen bromide peptide alpha 1(III)CB1,8,10,2
RT   (positions 223-402)."
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:821-832(1979).
RN   [3]
RP   SEQUENCE OF 423-571.
RX   MEDLINE=80026028; PubMed=488908;
RA   Bentz H., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. III. The
RT   amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4
RT   (positions 403-551)."
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).
RN   [4]
RP   SEQUENCE OF 572-808.
RX   MEDLINE=80026029; PubMed=488909;
RA   Lang H., Glanville R.W., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. IV. The amino
RT   acid sequence of the cyanogen bromide peptide alpha 1(III)CB5
RT   (positions 552-788)."
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).
RN   [5]
RP   SEQUENCE OF 809-947.
RX   MEDLINE=80026030; PubMed=488910;
RA   Dewes H., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. V. The amino
RT   acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A
RT   (position 789-927)."
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).
RN   [6]
RP   SEQUENCE OF 948-1049.
RX   MEDLINE=80026031; PubMed=488911;
RA   Allmann H., Fietzek P.P., Glanville R.W., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. VI. The amino
RT   acid sequence of the carboxyterminal cyanogen bromide peptide alpha
```

```
RT   1(III)CB9B (positions 928-1028).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).
CC   -!- FUNCTION: Collagen type III occurs in most soft connective tissues
CC       along with type I collagen.
CC   -!- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are
CC       linked to each other by interchain disulfide bonds. Trimers are
CC       also cross-linked via hydroxylysines.
CC   -!- PTM: Prolines at the third position of the tripeptide repeating
CC       unit (G-X-Y) are hydroxylated in some or all of the chains.
DR   PIR; A02862; CGBO7S.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01391; Collagen; 17.
DR   ProDom; PD000007; Clg_helix; 3.
DR   PROSITE; PS01208; VWFC_1; PARTIAL.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen.
FT   DOMAIN       1     14      NONHELICAL REGION (N-TERMINAL).
FT   DOMAIN      15   1040      TRIPLE-HELICAL REGION.
FT   DOMAIN    1041   1049      NONHELICAL REGION (C-TERMINAL).
FT   MOD_RES     95              HYDROXYLATION.
FT   MOD_RES    107              HYDROXYLATION.
FT   MOD_RES    119              HYDROXYLATION.
FT   MOD_RES    938              HYDROXYLATION.
FT   MOD_RES    950              HYDROXYLATION.
FT   CARBOHYD   107              O-LINKED (GAL. . .).
FT   CARBOHYD   950              O-LINKED (GAL. . .).
FT   DISULFID  1040   1040      INTERCHAIN.
FT   DISULFID  1041   1041      INTERCHAIN.
SQ   SEQUENCE 1049 AA;  93651 MW;  8EEC33D1C66EC9A3 CRC64;
```

```
Db   744 GKDGPPGPPGSNG 756
```

```
RESULT 15
CA13_HUMAN
ID   CA13_HUMAN     STANDARD;      PRT;   1466 AA.
AC   P02461; Q15112;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-OCT-2003 (Rel. 42, Last annotation update)
DE   Collagen alpha 1(III) chain precursor.
GN   COL3A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin fibroblast;
RX   MEDLINE=89350838; PubMed=2764886;
RA   Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,
RA   Prockop D.J.;
RT   "Structure of cDNA clones coding for the entire prepro alpha 1 (III)
RT   chain of human type III procollagen. Differences in protein structure
RT   from type I procollagen and conservation of codon preferences.";
RL   Biochem. J. 260:509-516(1989).
RN   [2]
RP   SEQUENCE OF 149-1225 FROM N.A.
RX   MEDLINE=89386015; PubMed=2780304;
RA   Janeczko R.A., Ramirez F.;
RT   "Nucleotide and amino acid sequences of the entire human alpha 1
RT   (III) collagen.";
RL   Nucleic Acids Res. 17:6742-6742(1989).
RN   [3]
RP   SEQUENCE OF 168-398.
RX   MEDLINE=77714724; PubMed=557335;
RA   Seyer J.M., Kang A.H.;
RT   "Covalent structure of collagen: amino acid sequence of cyanogen
RT   bromide peptides from the amino-terminal segment of type III collagen
RT   of human liver.";
RL   Biochemistry 16:1158-1164(1977).
RN   [4]
RP   REVISIONS.
RA   Seyer J.M.;
RL   Submitted (DEC-1977) to the PIR data bank.
RN   [5]
RP   SEQUENCE OF 399-727.
RX   MEDLINE=79000343; PubMed=687591;
RA   Seyer J.M., Kang A.H.;
RT   "Covalent structure of collagen: amino acid sequence of five
RT   consecutive CNBr peptides from type III collagen of human liver.";
RL   Biochemistry 17:3404-3411(1978).
RN   [6]
RP   SEQUENCE OF 728-964.
RX   MEDLINE=80198282; PubMed=6246925;
RA   Seyer J.M., Mainardi C., Kang A.H.;
RT   "Covalent structure of collagen: amino acid sequence of alpha 1
RT   (III)-CB5 from type III collagen of human liver.";
RL   Biochemistry 19:1583-1589(1980).
RN   [7]
RP   SEQUENCE OF 950-1466 FROM N.A.
RX   MEDLINE=88189827; PubMed=3357782;
RA   Mankoo B.S., Dalgleish R.;
RT   "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
RL   Nucleic Acids Res. 16:2337-2337(1988).
RN   [8]
RP   REVISION TO 1184.
RX   MEDLINE=89098346; PubMed=3211760;
RA   Molyneux K., Dalgleish R.;
RT   "Human type III collagen 'variant' is a cDNA cloning artefact.";
RL   Nucleic Acids Res. 16:11833-11833(1988).
RN   [9]
```

```
Query Match             62.1%;
Best Local Similarity   62.7%;
Matches  347;  Conservative   36;  Mismatches 161;  Indels   9;  Gaps   1;
                        Score 1905.5;  DB 1;  Length 1049;
                        Pred. No. 9.1e-71;

Qy    1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE   60
        |||  | |||||||  | ||  |  | |  ||| ||| |  ||  ||  || ||||  ||
Db  204 GSSGAPGQRGEPGPQGHAGAPGPGPPGSDGSPGGKGEMGPAGIPGAPGLIGARGPPGPP  263

Qy   61 GPGGPPGPKGDSGSPGEPGAPGSKGDTGAKGEPGPVVGVEGPPGPAGEGKPGARGPPGTGLP  120
        | |  |||  | |  | | | |  ||| || | |  | |  |   | ||  ||  ||  |||
Db  264 GTNGVPGGRGAAGEPGKNGAKGDPGPRGERGEAGSPGIAGPKGEDGKDGSPGEPGPGANGLP  323

Qy  121 GPPGERGERGPPGSRGFPGADGVAGPKGPAGERGSPGPAGEKGSPGEAGRPGEAGLPGAKGLT  180
        | || || ||  | |  ||  | ||  || |   |  | |  |  |  || ||   | | :::
Db  324 GAAGERGRVPGFRGPAGANGLPGEKGPPGDRGGFPGPGPAGPRGVAGPGRNGLPGGPGLRGIP  383

Qy  181 GSPGSPGDPGDKTGPPGPAGEDGRPGGPPGPGPPGEARGEAGVMGFPGPKGAAGERGVP  240
        |||| || ||| |||| |  ||||| | ||   ||   ||  | ||||||| | ||| |
Db  384 GSPGGPGSNGKPGPPGSQGETGRPGPPGSSPGPRGQPGVWGFPGPKGNDGAPGKNGERGGP  443

Qy  241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEEGPAGSPGFEGLPGPAGPPGEAGKPGSE  300
        | ||  | ||| | |  |||| |  || ||| | |  ||   |  ||  ||| ||| ||
Db  444 GGPGPQGPPAGKNGETGPQGPPGPTGPGSGDKGDTGPPGPPGQLQGLPGTSGPPGENGKPGEP  503

Qy  301 GVPGDLGAPGPSGARGEPGFPGPGERGVEGPPGPAGPPGPADGAPGDDGAKGDAGAPGAPGSE  360
        |  |  || |  | ||  | | |  |  ||| |  | |||  |||| |||| |||  | | |
Db  504 GPKGEAGADGIPGGKGDSGAPGERGPPGAGPPGPGGPPGPBGGKAAGPPGPPGPPGSA  563

Qy  361 GAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDK  420
        | ||| ||||||||  | | || |   || ||  | | |||| || |||||||| | |||
Db  564 GTPGLQGMPGERGGPGGPGPKGDKGEPGSSGVDGAPGKDGPRGPTGPIGPPGPAGQPGDK  623

Qy  421 GESGPSGPAGPTGARGAPGDRGEPGPPGPADGBEPGAKGEPGDAGAKGDAGPP  480
        ||||  | |  | |||  | | |  | ||  | |  ||  | ||  ||| |||
Db  624 GESGAPGVPGIAGPRGGPGRGEQPGPPAGPFPGAPGQNGEPGAKGERGAPGEKGEGGPP  683

Qy  481 GPAGPAGPGPIGDVGAPGAKGARGSAGPPGATGPFPGEAAGRVGPPGPGPSGDAGPP------  534
        |  | || |  | ||  | ||| | ||| |  |  || | ||  || | ||  ||
Db  684 GAAGPAGGSGPAGPPGPGGPGPGVKGERGSPGGPGAAGFPGGRGRPGPPGPGSNGNPGPPGSSGAP  743

Qy  535 --GPPGPAGKEG  544
```

RP   SEQUENCE OF 1065-1466 FROM N.A.
RX   MEDLINE=85087944; PubMed=6096827;
RA   Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
RA   Rosenbloom J., Myers J.C.;
RT   "Molecular cloning and carboxyl-propeptide analysis of human type III
RT   procollagen.";
RL   Nucleic Acids Res. 12:9383-9394(1984).
RN   [10]
RP   SEQUENCE OF 965-1200.
RX   MEDLINE=82008139; PubMed=7016180;
RA   Seyer J.M., Kang A.H.;
RT   "Covalent structure of collagen: amino acid sequence of alpha
RT   1(III)-CB9 from type III collagen of human liver.";
RL   Biochemistry 20:2621-2627(1981).
RN   [11]
RP   SEQUENCE OF 1176-1466 FROM N.A.
RX   MEDLINE=85157600; PubMed=2579949;
RA   Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;
RT   "Isolation of cDNA and genomic clones encoding human pro-alpha 1
RT   (III) collagen. Partial characterization of the 3' end region of the
RT   gene.";
RL   J. Biol. Chem. 260:4357-4363(1985).
RN   [12]
RP   SEQUENCE OF 1161-1200 FROM N.A.
RX   MEDLINE=86187804; PubMed=3754462;
RA   Miskulin M., Dalgleish R., Kluve-Beckerman B., Rennard S.I.,
RA   Tolstoshev P., Brantly M., Crystal R.G.;
RT   "Human type III collagen gene expression is coordinately modulated
RT   with the type I collagen genes during fibroblast growth.";
RL   Biochemistry 25:1408-1413(1986).
RN   [13]
RP   SEQUENCE OF 1-170 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=88303360; PubMed=3405773;
RA   Toman D., Ricca G., de Crombrugghe B.;
RT   "Nucleotide sequence of a cDNA coding for the amino-terminal region
RT   of human prepro alpha 1(III) collagen.";
RL   Nucleic Acids Res. 16:7201-7201(1988).
RN   [14]
RP   SEQUENCE OF 1-176 FROM N.A.
RX   MEDLINE=89378752; PubMed=2777083;
RA   Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;
RT   "Cloning and analysis of the 5' portion of the human type-III
RT   procollagen gene (COL3A1).";
RL   Gene 78:255-265(1989).
RN   [15]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97255959; PubMed=9101290;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT   associated collagen (type IX), and network-forming collagen (type X)
RT   cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL   Hum. Mutat. 9:300-315(1997).
RN   [16]
RP   VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.
RX   MEDLINE=93293988; PubMed=8514866;
RA   Tromp G., Wu Y., Prockop D.J., Madhatheri S.L., Kleinert C.,
RA   Earley J.J., Zhuang J., Noerrgaard O., Darling R.C., Abbott W.M.,
RA   Cole C.W., Jaakkola P., Ryynaenen M., Pearce W.H., Yao J.S.T.,
RA   Majamaa K., Smullens S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,
RA   Jackson C.E., Michels V.V., Kaye M., Kuivaniemi H.;
RT   "Sequencing of cDNA from 50 unrelated patients reveals that mutations
RT   in the triple-helical domain of type III procollagen are an
RT   infrequent cause of aortic aneurysms.";
RL   J. Clin. Invest. 91:2539-2545(1993).
RN   [17]
RP   VARIANT THR-698.
RX   MEDLINE=91045136; PubMed=2235526;
RA   Zafarullah K., Kleinert C., Tromp G., Kuivaniemi H., Kontusaari S.,
RA   Wu Y., Ganguly A., Prockop D.J.;
RT   "G to A polymorphism in exon 31 of the COL3A1 gene.";
RL   Nucleic Acids Res. 18:6180-6180(1990).
RN   [18]

RP   VARIANT AORTIC ANEURYSM ARG-786.
RX   MEDLINE=91056145; PubMed=2243125;
RA   Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;
RT   "A mutation in the gene for type III procollagen (COL3A1) in a family
RT   with aortic aneurysm.";
RL   J. Clin. Invest. 86:1465-1473(1990).
RN   [19]
RP   VARIANT EDS-IV ARG-828.
RX   MEDLINE=94016385; PubMed=8411057;
RA   Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;
RT   "The substitution of glycine 661 by arginine in type III collagen
RT   produces mutant molecules with different thermal stabilities and
RT   causes Ehlers-Danlos syndrome type IV.";
RL   J. Med. Genet. 30:690-693(1993).
RN   [20]
RP   VARIANT EDS-IV SER-957.
RX   MEDLINE=91109135; PubMed=2492273;
RA   Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;
RT   "A single base mutation that substitutes serine for glycine 790 of
RT   the alpha 1 (III) chain of type III procollagen exposes an arginine
RT   and causes Ehlers-Danlos syndrome type IV.";
RL   J. Biol. Chem. 264:1349-1352(1989).
RN   [21]
RP   VARIANT EDS-IV VAL-960.
RX   MEDLINE=95268429; PubMed=7749417;
RA   Tromp G., de Paepe A., Nuytinck L., Madhatheri S.L., Kuivaniemi H.;
RT   "Substitution of valine for glycine 793 in type III procollagen in
RT   Ehlers-Danlos syndrome type IV.";
RL   Hum. Mutat. 5:179-181(1995).
RN   [22]
RP   VARIANT EDS-IV GLU-1014.
RX   MEDLINE=92316511; PubMed=1352273;
RA   Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,
RA   Pope F.M.;
RT   "A single base mutation in the gene for type III collagen (COL3A1)
RT   converts glycine 847 to glutamic acid in a family with Ehlers-Danlos
RT   syndrome type IV. An unaffected family member is mosaic for the
RT   mutation.";
RL   Hum. Genet. 89:414-418(1992).
RN   [23]
RP   VARIANT EDS-IV ASP-1050.
RX   MEDLINE=90037070; PubMed=2808425;
RA   Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;
RT   "Single base mutation in the type III procollagen gene that converts
RT   the codon for glycine 883 to aspartate in a mild variant of
RT   Ehlers-Danlos syndrome IV.";
RL   J. Biol. Chem. 264:19313-19317(1989).
RN   [24]
RP   VARIANT EDS-IV VAL-1077.
RX   MEDLINE=91374480; PubMed=1895316;
RA   Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Narcisi P.,
RA   Richards A.J., Lloyd J.C., Ward P.N., de Paepe A., Narcisi P.,
RA   Pope F.M.;
RT   "Characterisation of a glycine to valine substitution at amino acid
RT   position 910 of the triple helical region of type III collagen in a
RT   patient with Ehlers-Danlos syndrome type IV.";
RL   J. Med. Genet. 28:458-463(1991).
RN   [25]
RP   VARIANT EDS-IV GLU-1173.
RX   MEDLINE=93022543; PubMed=1357232;
RA   Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;

Query Match   61.9%; Score 1901; DB 1; Length 1466;
Best Local Similarity   63.1%; Pred. No. 1.7e-70;
Matches 345; Conservative 37; Mismatches 159; Indels 6; Gaps 2;

```
Qy    1 GSEGPEGVRGEPGPPGPAGAAGPAGDPGADGEPGAKGADGAPGIAGAPGFPGARGPSGPE 60
          ||||||| || ||||| || ||||  | ||||||||| ||||||||  ||| |
Db  360 GSNGAPGQRGEPGPGHAGAQGPGPPGINGSPGGKGEMGPAGIPGADPGLMGARGGPPGPA 419

Qy   61 GPGPPGKGDSGEPGAPGSKGDTGAKGEPGPGVGVEGPPGPPGAGEBGKPGGARGEPGPTGLP 120
          |||| | | |:|| ||:| |||::|: |:||||:|||||||||||
Db  420 GANGAPGLRGGAGEPGKNGAKGEPGPRGRGERGEAGIPGVPGAKGEDGKDGSPGEPGANGLP 479
```

Fri May 7 16:57:55 2004

us-10-658-989a-4.rsp

```
Qy   121 GPPGERGPGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
             |||   ||| |   ||:|: ||| |||| ||: | | ||||: ||   | |  :: ||::
Db   480 --GAAGERGAPGFRGPAGPNGIPGEKGPAGERGAPGPAGPRGAAGEPGRDGVFGGPGMRGMP 539

Qy   181 GSPGSPGPDGKTGPPGPAGEDGDRPGPPGPPGPPGAARGEAGVMGFPGPKGAAGEPGKAGERGVP 240
             |||||  || |  || || :  |||:  |  :|||  |||||  || |||:|||
Db   540 GSPGGPGSDGKPGPPGSQGESGRPGPPGPGPGPPGQPGVMGFPGPKGNDGAPGKNGERGGP 599

Qy   241 GPPGAVGPAGKDGEAGAEGPPGPAGPAGERGEBGPAGSPGFEGLPGPAGPPGEAGKPGEE 300
             |   ||  ||  |||: |: |||  |:||:: ||| | |||| | |||| ||||
Db   600 GGPGPQGPPGKNGETGPQGPPGPTGPGGDKGDTGPPGPQGLQGLPGTGGPPGENGKPGEP 659

Qy   301 GVPGDLGAPGPSGARGEPGFPGERGVEGPPGPAGPP---GADGAPGDDGAKGDAGAPGAP 357
             |  :  |||| |||||||  |   |||  ||||   | :: ||||  :|   ||||
Db   660 GPKGDAGAPGKGDAGAPGER---GPPGLAGAPGLRGGAGPPGEBGGKGAAGPPGPP 716

Qy   358 GSEGAPGLEGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAP 417
             |:  | ||||||| |:||| ||||||||||:
Db   717 GAAGTPGLQGMPGERGGLGSPGPKGDKGBPGGPGADGVPGKDGPRGPTGPIGPPGPAGQP 776

Qy   418 GDKGESGPSGPAGPTGARGAPGDRGEPGPPGPAGFAGPPGADGEPGDAGAKGDA 477
             ||:| ||  | ||||||||||| | || : | |:: :| :||: ||
Db   777 GDKGEGGAPGLPGIAGPRGSPGERGETGPPGPAGFPGAPGQNGEPGGKGERGAPGEKGEG 836

Qy   478 GPPGPAGPAGPPGPIGDVGAPGAKGARGSAGPPGATGPGAAGRVGPPGPSGDAGPPGPP 537
             |||| | || :|||  |: |  | |:||| ||| | :|: |||| |||
Db   837 GPPGVAGPPGGGSGPPGPQGVKGERGSPGGGFPGAAGFPGARGLPGGPGGSNGNPGPPGPS 896

Qy   538 GPAGKEG 544
             | |:|:
Db   897 GSPGKDG 903
```

Search completed: May 7, 2004, 16:28:49
Job time : 19 secs